2016-06

IED pattern recognition using sinusoidal models

Ardohain, Christopher M.

Monterey, California: Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

IED PATTERN RECOGNITION USING SINUSOIDAL MODELS

by

Christopher M. Ardohain

June 2016

Thesis Advisor: Michael Atkinson
Second Reader: Connor McLemore

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704–0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704–0188) Washington DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2016	**3. REPORT TYPE AND DATES COVERED** Master's thesis	
4. TITLE AND SUBTITLE IED PATTERN RECOGNITION USING SINUSOIDAL MODELS			**5. FUNDING NUMBERS**
6. AUTHOR(S) Christopher M. Ardohain			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943–5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number____N/A____.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

More than half of all U.S. casualties in Iraq and Afghanistan were caused by improvised explosive devices (IEDs). Despite the spending of over $75 billion to combat this threat, intelligence analysts still lack efficient tools to conduct IED pattern analysis. This thesis evaluates sinusoidal models for effectiveness in assisting in the identification of IED patterns.

We formulate three models to test against IED patterns encountered in Iraq and Afghanistan: the Hawkes point process, the non-linear optimization of a sine function, and discrete Fourier transforms (DFT). Non-linear optimization and DFT models both out-perform a mean inter-arrival model when applied to representative IED patterns. We also applied these models against portions of an Iraq IED dataset using a rolling horizon forecast. Lastly, we test model performance when applied to patterns identified from the Iraq dataset. We conclude that although there is not a "silver bullet" for IED pattern detection, the use of these models in IED environments has the potential to reduce the amount of time and effort intelligence analysts expend when identifying IED patterns. We recommend incorporating these models into a graphic user interface usable by intelligence analysts responsible for IED pattern recognition.

14. SUBJECT TERMS improvised explosive device, IED, Hawkes point process, discrete Fourier transforms, pattern recognition, rolling horizon forecast			**15. NUMBER OF PAGES** 123
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540–01–280–5500 Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

IED PATTERN RECOGNITION USING SINUSOIDAL MODELS

Christopher M. Ardohain
Major, United States Army
B.S., United States Military Academy, 2005

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2016

Approved by: Michael Atkinson
Thesis Advisor

Connor McLemore
Second Reader

Patricia Jacobs
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

More than half of all U.S. casualties in Iraq and Afghanistan were caused by improvised explosive devices (IEDs). Despite the spending of over $75 billion to combat this threat, intelligence analysts still lack efficient tools to conduct IED pattern analysis. This thesis evaluates sinusoidal models for effectiveness in assisting in the identification of IED patterns.

We formulate three models to test against IED patterns encountered in Iraq and Afghanistan: the Hawkes point process, the non-linear optimization of a sine function, and discrete Fourier transforms (DFT). Non-linear optimization and DFT models both outperform a mean inter-arrival model when applied to representative IED patterns. We also applied these models against portions of an Iraq IED dataset using a rolling horizon forecast. Lastly, we test model performance when applied to patterns identified from the Iraq dataset. We conclude that although there is not a “silver bullet” for IED pattern detection, the use of these models in IED environments has the potential to reduce the amount of time and effort intelligence analysts expend when identifying IED patterns. We recommend incorporating these models into a graphic user interface usable by intelligence analysts responsible for IED pattern recognition.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ARIMA	Autoregressive Integrated Moving Average
ASR	alternate supply routes
CENTCOM	Central Command
CIDNE	Combined Information Data Network Exchange
CPAT	Crime Pattern Analysis Team
DFT	Discrete Fourier Transform
DoD	Department of Defense
FOIA	Freedom of Information Act
IED	Improvised Explosive Device
JIDA	Joint Improvised-Defeat Agency
JIEDDO	Joint IED Defeat Organization
LP	linear program
MLE	Maximum Likelihood Estimation
MSR	main supply routes
NAI	named area of interest
NHPP	Non-Homogenous Poisson Process
NL	non-linear
RCT	Route Clearance Team
RMSE	Root Mean-Squared Error
SEPP	Self-Exciting Point Process
SIGACTs	Significant Activities
USD	United States Dollars

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The insurgent weapon of choice in Iraq and Afghanistan is the improvised explosive device (IED). According to Gregg Zoroya of *USA Today,* this crude but effective weapon has caused over 50 percent of U.S. casualties and prompted the U.S. to spend over $75 billion on new vehicles, armor, and other detection/defeat technologies. A report written by the Action on Armed Violence spanning 66 countries shows a nearly 70 percent increase in civilian casualties due to IEDs from 2011 to 2013. IED statistics suggest that IEDs will remain a weapon of choice to target conventional U.S. forces and terrorize civilian populations. The objective of this thesis is to improve tactical level targeting of IEDs by developing models to assist intelligence analysts conducting IED pattern recognition.

While researchers have developed mathematical models for examining IED activity, our approach attempts to identify IED patterns that would allow an intelligence analyst to specify a likely date, time and location for follow-on IED events. IED patterns are most easily identifiable, both spatially and temporally, at a local level. It is therefore necessary to filter the data to only the most recent events (we use fewer than 25) in a small area (we do not exceed 40km of road). Pattern identification is difficult; most identifiable patterns in Iraq and Afghanistan were the result of the cyclical nature of IED supply or coalition force patterns in the area. Analysts identified patterns by examining the timing of events in the area of interest. Specifically, analysts focused on the time between successive events (i.e., the *inter-arrival time*). As an example, a pattern may consist of three to six IED events in quick succession followed by one to two events with longer inter-arrivals. The underlying basis for this type of pattern may be an IED supply cycle where completed IEDs are available in large batches and emplaced in quick succession. The long inter-arrivals may be indicative of spreading out emplacements while awaiting the next IED batch. These patterns create distinct

sinusoidal shapes when the inter-arrival times are plotted against observation numbers.

Our objective is to identify a model that could clearly distinguish a random sequence of inter-arrivals from legitimate patterns. We formulate three models to fit a sequence of inter-arrival times. These include the Hawkes point process, the non-linear (NL) optimization of a sine function, and discrete Fourier transforms (DFT). We artificially constructed ideal versions of several common pattern types to use as test data during model development. We also randomly generated a sequence to provide a comparison. We measure model performance using root mean-squared error (RMSE), comparing model predictions to the actual data. To evaluate the model in this phase, we compute the RMSE using the same observations that generated model parameter estimates.

We first test a Hawkes point process model. The Hawkes process is an arrival process where one arrival can trigger future arrivals, similar to the aftershocks of an earthquake. One IED event may trigger future IED events. The Hawkes process performs poorly, however. We believe the primary reason for this is that the underlying dataset is small. Hawkes point process is typically applied to large datasets. We then use NL optimization to fit a sine function to the data, because of the sinusoidal or cyclical nature of common IED patterns. NL functions are computationally expensive and therefore are not suitable for situations where we have to evaluate thousands of sequences. Our last approach uses DFT, which represents data as a linear combination of sine and cosine functions, similar to a polynomial function. Closed formed expressions exist to fit the DFT parameters, which makes the computational run times trivial. DFT is more suitable than NL optimization when the number of sequences being evaluated becomes very large. The NL optimization model and DFT models are unable to distinguish idealized patterns from random data.

To better distinguish patterns from randomness, we develop a methodology, we call “test-two,” to fit observations outside of a test sample. If our sequence has N observations, we fit our models using the first N-2 inter-arrivals

and then we predict the *N*-1 and *Nth* inter-arrival. We calculate the RMSE using only the last two predicted observations. The results of our test-two methodology suggest there may be some potential to utilize NL optimization and DFT models to distinguish between random sequences of IEDs and patterns.

We apply the NL optimization and DFT models using test-two methodology to a real-world Iraq dataset with a rolling horizon forecast. The dataset consists of all recorded, real-world Iraq IED events from January, 2005 to December 2008, over 80,000 IED events in total. We subset the Iraq data into eight, 6-month windows and filter the data spatially, which results in a total of 1328 IED sequences to test. The rolling horizon forecast provides a quick way to compare the performance of our models against a naïve model that assumes all inter-arrivals are the mean as determined by the sample. For each subset, we produce a RMSE for all three models (DFT, NL optimization, naïve) for the first set of observations. We then step forward one observation and repeat the process; the second sequence differs from the first sequence by one observation. The naïve model out-performed the other models. This is not surprising since the naïve model blindly plows through all 1328 sequences, even though actual patterns are rare and most of the 1328 sequences do not constitute legitimate patterns. The DFT model performs better than the naïve approximately one-third of the time. The NL optimization model performs similarly. Applying the DFT and NL optimization across all sequences in a brute force manner is not effective: we generate better aggregate results by using the naïve inter-arrival. Rather than apply our methods across all sequences, we then applied our models to sequences in the Iraq dataset visually identified as patterns.

Using the same 1328 IED sequences, we visually inspect each sequence manually and find 19 pattern candidate sequences that would warrant future targeting. Of the 19 candidate patterns, we find that the naïve model *never* performs the best and that the NL optimization and DFT model perform better than the mean approximately 85 percent of the time. For this particular subset of

data, the results suggest that an analyst could use the NL optimization model to filter down the number of sequences requiring visual inspection. Specifically, the filter reduces the number of sequences to inspect from 1328 to 440 and this smaller subset contains 90 percent of the visually identified patterns.

We conclude that although there is not a “silver bullet” to IED pattern detection, making these models available to IED analysts has the potential to greatly reduce the amount of time and effort intelligence analysts expended to identify IED patterns. We recommend incorporating these models into a graphic user interface usable by intelligence analysts responsible for IED pattern recognition.

References

AOAV (2014) The impact of IEDs: Three years of data, 2011–2013. *Action on Armed Violence* (December 1), https://aoav.org.uk/2014/three-years-ieds/.

Zoroya, G (2013) How the IED changed the U.S. military. *USA Today* (December 19), http://www.usatoday.com/story/news/nation/2013/12/18/ied-10-years-blast-wounds-amputations/3803017/.

ACKNOWLEDGMENTS

I would like to express gratitude to several individuals who were pivotal to the completion of this thesis. They include all those who helped guide me as a student, a career Army officer, and as an individual.

First, I am incredibly indebted to my thesis advisor, Dr. Michael Atkinson, who was willing to work with a student passionate about a project, despite the fact that it was not sponsor driven. Your countless hours of guidance allowed me to explore many methodologies previously unknown to me. Also, I thank you for your tireless efforts to edit and correct my work throughout this process. Thank you for your wisdom and patience.

To my Army mentors including COL David Theisen, COL Robert Whittle, COL Jim Koeppen, LTC Tom Hairgrove, and LTC Jason Williams, to name a few during my time with the “Best Engineer Battalion Ever Formed”—thank you all so much for everything you taught me. I doubt I would be the officer I am today without your push to make me better. I would also like to thank and recognize the extraordinary work and effort contributed by the Soldiers assigned to my Intelligence sections. It will always be remembered.

I would also like to thank my parents, Steve and Marian, who taught me and truly believed that with hard work and a little luck, I could achieve anything I wanted. Most importantly, I would like to thank my wife, Kate, and son, Steven, for their love and support throughout the inception and creation of this thesis. You are the bedrock upon which this stands and I am greatly looking forward to making up all the time that I missed. I love you both.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. MOTIVATION

Improvised Explosive Devices (IEDs) originated in the 1500s in the form of ships laden with explosive materials (Singer 2012). The term "IED" today is synonymous with fixed explosives detonated remotely, such as those used against U.S. military vehicles and personnel in Iraq and Afghanistan. IEDs are the weapon of choice for insurgencies due to their low cost, minimal risk to insurgents, and their effectiveness against their intended targets. At their peak, insurgents emplaced over 2,700 IEDs a month in Iraq alone (Atkinson 2007) resulting in significant U.S., Coalition and civilian casualties. IEDs became the weapon of choice in both Iraq and Afghanistan even though there were significant differences in targets chosen, explosives used, and emplacement methods between the two countries. Insurgents found an effective, asymmetric way to inflict great pain on a vastly superior conventional force using a few munitions and a detonator held together with electrical tape.

IEDs have caused over half of all U.S. casualties in Iraq and Afghanistan (Zoroya 2013), making IEDs the single most effective weapon system employed by the insurgents. This has prompted the United States and its allies to spend over $75 billion on new vehicles, armor, and other detection/defeat technologies (Zoroya 2013). At an average cost of $265 USD per IED (Ackerman 2011), IEDs have proved to be extremely cost effective. As the U.S. and its allies look toward potential future conflicts, there is a consensus that enemies will employ IEDs in the future given their success in Iraq and Afghanistan. A report written by the Action on Armed Violence shows a nearly 70 percent increase in civilian casualties due to IEDs from 2011 to 2013 (AOAV 2013). IED reports in that study span 66 countries and are not exclusive to the wars in Iraq and Afghanistan. IED statistics suggest that IEDs will remain a weapon of choice to combat superior military force, and terrorize civilian populations. The objective of this thesis is to

improve tactical level targeting of the IEDs by developing models to assist intelligence analysts conducting IED pattern recognition.

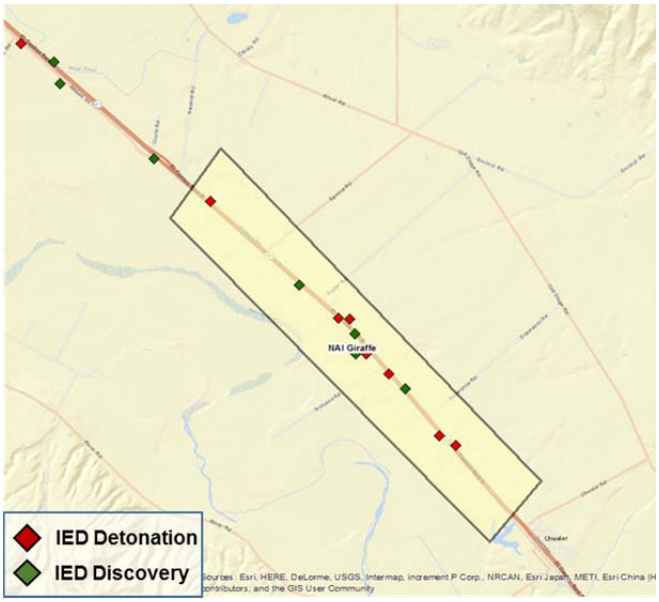
B. BACKGROUND

Part of the response to IEDs included the development of entire organizations dedicated to the problem. On February 14, 2006, the Joint IED Defeat Organization (JIEDDO) was established under DoD directive (JIDA 2015). As the lead agency in combating the IED threat, JIEDDO (now known as the Joint Improvised-Defeat Agency or JIDA) developed a strategy that includes *Attack the Network* and *Defeat the Device* (JIDA 2015). *Attack the Network* consists of tactical to operational level targeting of the network of individuals connected to the production, distribution, and emplacement of IEDs. These individuals include, but are not limited to, financiers, bomb-makers, and cell leaders. There is a generalized belief that if a node can be neutralized high enough in the network, the effects will trickle down with the potential to prevent multiple IED emplacements. *Defeat the Device* consists of developing technologies to detect and defend against individual IED attacks. Although JIDA developed and produced techniques and equipment for the entire force, it heavily focused on Route Clearance Teams (RCTs).

RCTs were typically composed of Army Engineers using specialized equipment to detect and disable IEDs along main supply routes (MSRs) and alternate supply routes (ASRs). The primary objective of RCTs was the interdiction of the device before it could be detonated against a target. Once an RCT discovered an IED, the team's use of robots and articulating arms mounted on vehicles allowed the interrogation of suspicious sites with limited exposure to individuals. RCTs were also better equipped and armored to potentially survive an IED detonation than their conventional or logistical force counterparts. JIDA developed, tested, and sourced many of these technologies. Anecdotal evidence suggests that the presence of Route Clearance had no significant impact on the number of IEDs emplaced by insurgents; however, RCTs did prove effective in

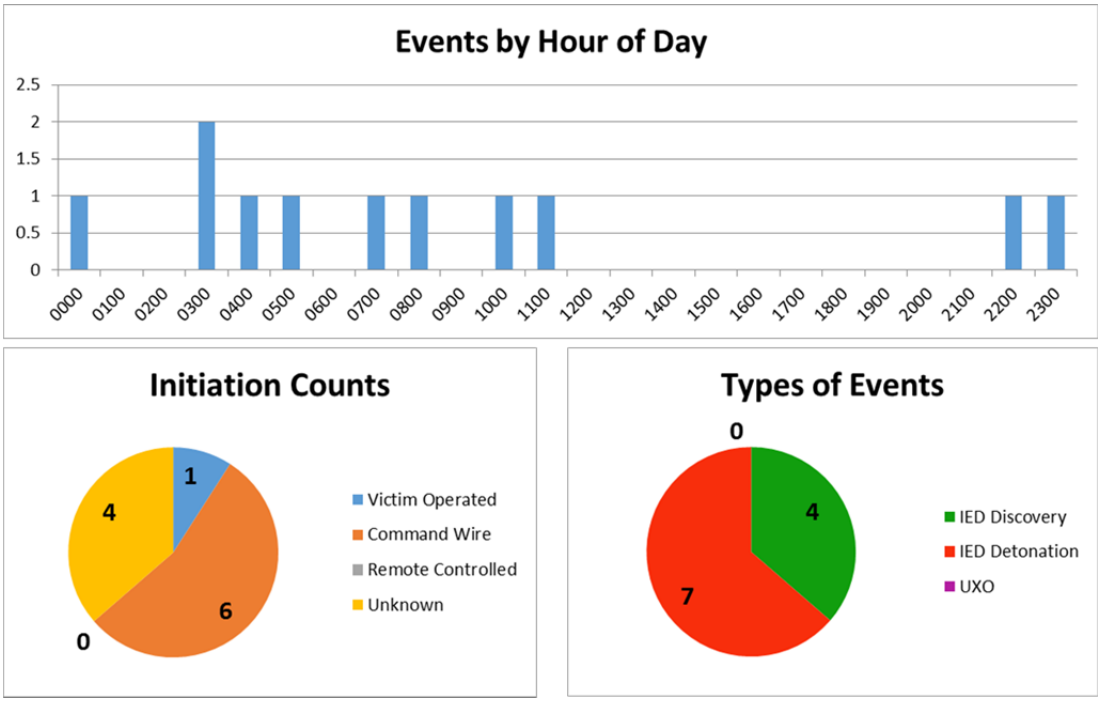
terms of higher discovery rates and lower casualties than maneuver or logistic forces. This is in part due to specialized equipment, which allowed distanced interrogation of possible IEDs, and perhaps more importantly, to the repetitive nature of Route Clearance. Route Clearance Soldiers would typically perform one eight- to 12-hour mission per day, six days a week, with one day for maintenance and recovery. During these missions, the same routes were consistently cleared, which allowed the development of an intense familiarity with the environment, enabling detection of small changes possibly indicating recent suspicious human activity.

In support of RCTs, Military Intelligence Soldiers provided analysis to maximize the IED involvement of the RCTs. In my experience, most of this analysis was rather unsophisticated. The analysts created a named area of interest (NAI) around a region with a higher preponderance of IED events (Figure 1). The analysts rarely updated the spatial boundaries of the NAIs (e.g., once every three to six months) because of difficulties in sharing changes between all interested units. Additionally, the analysis of the IED events within a single NAI tended to focus on summary statistics. In most cases, the analysis would inform the decision maker about the most likely time of day, initiation types, discovery/detonation rates, types of targets, etc. based on the historical events of the past week or month (Figure 2). This type of analysis was not unique to Route Clearance units, however, as logistics units were also very concerned about IED threats on supply routes. Even Intelligence Analysts supporting maneuver forces briefed their patrols on likely locations and times for IEDs prior to every mission and would use similar summary statistics to describe the IED environment.



Analysts would capture the spatial boundaries of high activity using NAIs and would use those boundaries as the basis for statistical analysis.

Figure 1. Example of a Named Area of Interest (NAI).



Replicated analysis highlighting the statistical history of a particular NAI. Graphics like these would inform a commander about likely IED event times, types of initiation, and the discovery vs. detonation rate for that area.

Figure 2. Traditional Statistical Analysis by NAI.

Presentations of summary statistics became the standard in Iraq and Afghanistan. As a deployed intelligence analyst, I found this type of basic analysis adequate to describe past IED activity, but it suffered from numerous flaws. The environment was described on a sliding scale of probability (higher probability of IED activity from this time to this time in this location). This was informative to planners who were creating base route clearance schedules, but it failed to provide accurate predictions about specific locations and times for future events that would prompt an allocation of surveillance or clearance assets. Another issue with this type of analysis was the use of NAIs to spatially subset IED activity. Although the intended purpose of using NAIs was to focus analysis on areas of high IED activity, they became a visual obstacle preventing intelligence analysts from exploring the influence of external IED activity. Military Intelligence Soldiers supporting route clearance and logistics units defined success as being able to identify specific IED patterns to “predict” future events. Usually, summary statistics based on NAIs did not provide the necessary information to meet this need.

In my experience, intelligence analysts saw the most success when they analyzed the time between IED events for a given geographic area rather than focusing singularly on event time of day. On rare occasions, they could even identify patterns in the inter-arrival times that could be exploit to predict future IED events or series of events. Analysts found that, on occasion, enemy activity naturally fell into an identifiable pattern, which was sometimes based on or influenced by IED supply cycles, insurgent daily life, or patterns set by U.S. forces. Regrettably, this type of analysis was painstaking; an analyst would manually evaluate potentially hundreds of IED sequences searching for patterns. Analysis becomes even more complicated when one considers the possibility of patterns only existing based on categorical variables such as initiation type.

C. SCOPE

Proper employment of route clearance assets in space and time requires military planners to consider basing, routes to patrol, and patrol times that maximize IED involvement rates. These three topics have been researched thoroughly over the past decade, but a gap still exists for efficiently recognizing patterns given previous IED activity in well-defined space. Successful IED pattern recognition would prompt a decision to alter a patrol schedule or request a surveillance asset in order to interdict the emplacement of the next device

I spent 24 months facing this problem without discovering an efficient method to evaluate potential patterns. This thesis provides a more rigorous methodology for IED intelligence analysis that streamlines the processes and provides better results faster than current manual approaches. We evaluate various mathematical methods to determine the predictability of IED sequences given a defined geographic area. We assume that the spatial boundaries will be selected by the intelligence analyst each time this analysis is performed; we focus on temporal pattern identification using time between IED events as the only model input. This thesis will not take into account external factors that often disrupt IED patterns such as the death or capture of insurgent leaders or the departure of Coalition forces from a particular area. Rather, it will focus on identifying possible patterns for future exploitation by U.S. and coalition forces.

D. LITERATURE REVIEW

Numerous researchers have developed mathematical models to examine IED activity. These models include stochastic Markov chains, game theory, and point processes. In fact, JIDA employs a team called the Crime Pattern Analysis Team (CPAT), which consists of mathematicians and law enforcement experts who are responsible for developing and implementing predictive IED models (Shankar 2014). Those tools provide predictive IED analysis to deployed commanders. Unfortunately, CPAT is limited by its small size, which limits the amount of information and data it can process at any given time. CPAT products

are also often enhanced by biometrics obtained during exploitation of an IED event. This again adds robustness to its analysis but at the cost of time. These limitations force the CPAT to concentrate on requests from deployed units on a first come, first served basis. CPAT has neither the capability to recommend route clearance schedules to all commanders across a theatre nor the capability to recommend a geographic target for an ad hoc surveillance asset transiting battlespace. The methodologies we propose would not only assist intelligence officers supporting deployed units, but have the potential to assist subject matter experts like those at CPAT by quickly identifying potential patterns for further analysis.

Rather than attempting to identify IED patterns, one analytic approach is to optimize route clearance scheduling based on probability modeling. This method was explored extensively by LTC Christopher Marks in 2009. Using an effectiveness parameter and modeling IED arrivals as a Poisson process, Marks' algorithm allows the user to define IED risk along a section of a route (Marks 2009). Route Clearance response to that risk is then optimized using mixed integer linear programming with the route network and availability of RCTs as constraints. Marks highlights that possible future work in this field could include more in-depth modeling of enemy activity to more accurately feed his route clearance optimization (Marks 2009). He specifically suggests the possible use of game theory models, such as those developed by Alan Washburn, as potential methods to add robustness to enemy modeling (Washburn, 2006).

Similar work by Professor Robert Koyak from the Naval Postgraduate School determines the probability of encountering an IED along a particular stretch of road based on not only previous activity but also friendly force traffic patterns (Koyak 2009). Koyak begins with the assumption that IED encounters occur no later than the next passing friendly convoy, which allows him to assert that the emplacement of that IED occurred sometime after the last friendly presence along that particular stretch of road. Using this idea as a framework, he extends the model to allow for specific enemy targeting and the possibility of IED

discovery if the convoy is not the intended target (Koyak 2009). His work in conjunction with the route clearance optimization developed by Marks could be developed into a powerful tool for developing a base route clearance schedule. However, a gap still remains as neither method identifies IED patterns that would lead an intelligence analyst to recommend a change in schedule or request surveillance of a specific location at a specific time.

Matthew Bengini and Reinhard Furrer worked jointly on an IED prediction method that used spatial and temporal clustering in an attempt to optimize surveillance loiter time and location (Bengini et al. 2012). They created an intensity function based on IED event time of day and distance away from a known origin, which then can be transformed into a contour map of probability densities. The optimal assignment of surveillance assets is then calculated based on maximizing the integral of probabilities for a given length of time and width of search range. They discovered that this method enjoyed some success only in the short term. However, this methodology was developed to inform a planner constructing a base surveillance schedule and it does not identify specific patterns based on time between events (Bengini et al. 2012).

The most recent academic work on IED activity was conducted in 2014. MAJ Arun Shankar produced a dissertation attempting to model IED activity, specifically in Afghanistan against dismounted patrols (Shankar 2014). This problem set is uniquely different from IED activity against mounted patrols due to the possibility of movement that deviates from an established route or road network. The ability to move in two-dimensional space as opposed to a one-dimensional road (forward or backward) creates an environment where there is a much greater probability that emplaced IEDs were missed by passing troops. Through his work, Shankar developed three different models to describe and predict future IED activity. He evaluated IED activity temporally and rejected the hypothesis that IED activity targeting dismounted patrols can be described by a non-homogenous Poisson process (NHPP). He developed a spatial clustering model that simulates future IED events using historical data. Lastly, he focused

on modeling emplacement time rather than IED discovery or detonation times using overlapping patrol zones and probabilities of each patrol encountering a single IED. He found that although the IED events could not be described as NHPP, their emplacement could. He points out that NHPPs rely on large datasets to be validated which help inform operational commanders about upward or downward IED trends across an entire theatre (Shankar 2014). Shankar's work, like so many other models of IED activity, has not resulted in a tangible tool or method usable by IED analysts.

Given the periodic and spatial nature of IED activity, a similar field that has received attention from analysts is crime patterning. Crime analysis is a long-established practice; however, it was not recognized as a profession and did not receive a name until the early 1970s (Stevenson 2013). One often-used method for forecasting future crime patterns is time series analysis. Time series analysis is especially useful when attempting to predict patterns with seasonality using occurrences of previous events to build a prediction of the future. These types of models are often referred to as Autoregressive Integrated Moving Average (ARIMA) and have successfully been used to model crime. Esra Polat created one such model by clustering crime activity spatially and applying the Box-Jenkins ARIMA model (Polat 2007). One significant disadvantage to time series analysis is the need for large data due to the need to capture seasonality. Polat's model attempted to predict activity on a daily basis over the course of a year, and he therefore required three years of data to sufficiently describe seasonality (Polat 2007). Anecdotal evidence suggests that IED pattern recognition is most effective when IED data is filtered to a very local level, both spatially and temporally, which results in very sparse data unsuited for the modeling techniques suggested by Polat.

Crime patterning is also explained using various self-exciting point process (Mohler et al. 2012). While a traditional point process has a single arrival rate, a self-exciting point process is defined by a variable arrival rate triggered by additional arrivals into a system. A classic application a self-exciting point

process model is seismic activity where one earthquake can trigger multiple aftershocks (Ogata 2012). Mohler cites several examples where a self-exciting point process would be appropriate to model arrival behavior. Those examples include gang violence where a single gang shooting triggers retaliatory violence and home robberies where burglars often case several houses in a neighborhood before attempting their first robbery and subsequent robberies happen in quick succession. The authors use large datasets to explore the self-exciting point process as a model of crime. It is possible that cyclical IED patterns can be described by a similar process when re-supply cycles are contributing factors to IED arrival rates at a local level.

In 1999, Dan Helms wrote a technical report entitled “The Use of Dynamic Spatio-Temporal Analytic Techniques to Resolve Emergent Crime Series” (Helms 1999). In this report, Helms details various techniques to predict and interrupt crime patterns. One method he details is the use of hydrology mapping to describe the temporal nature of crime. It consists of mapping activity with day of week as one axis and time of day as another. Each crime represents added weight, creating hydrology contours which could then be used to refine predictions of future events.

Another method described by Helms was the use of Discrete Fourier Transforms for pattern recognition. Using the most influential harmonics from spectral analysis, an analyst can estimate the periodicity and therefore develop a prediction for future events based on time between crimes (Helms 1999). This particular method shows promise for modeling IED activity since localized patterns are often cyclical in nature. This is most likely the result of either an IED supply cycle constraining the number of IEDs that can be emplaced at any given time or caused by the cyclic presence of coalition forces in the area. Fourier Transforms are therefore of particular interest in this thesis and will be explored in much greater detail in subsequent chapters.

E. THESIS STRUCTURE

The remainder of this thesis is organized into four additional chapters. Chapter II describes the two sources of data available for this thesis as well as highlights patterns based on time between IED events. The following chapter will test three different methods for pattern prediction against known patterns: Hawkes point process, non-linear optimization of sinusoidal functions, and discrete Fourier transforms. Chapter IV will explore the methodology of using a test set to confirm model results as well as performing rolling predictions on IED events in Iraq. The last chapter will provide a summary of the thesis findings and potential future work.

THIS PAGE INTENTIONALLY LEFT BLANK

II. DATA COLLECTION AND FILTERING

This chapter will provide an overview of the two data sources used to build and test our models. We provide the summary statistics for each and explore IED emplacements as a Poisson process. We then describe the types of IED patterns common in Iraq and Afghanistan. We also discuss methodologies for data filtering since we want to identify IED patterns at a local level. The chapter concludes with a discussion about the necessary requirements to distinguish patterns from random data.

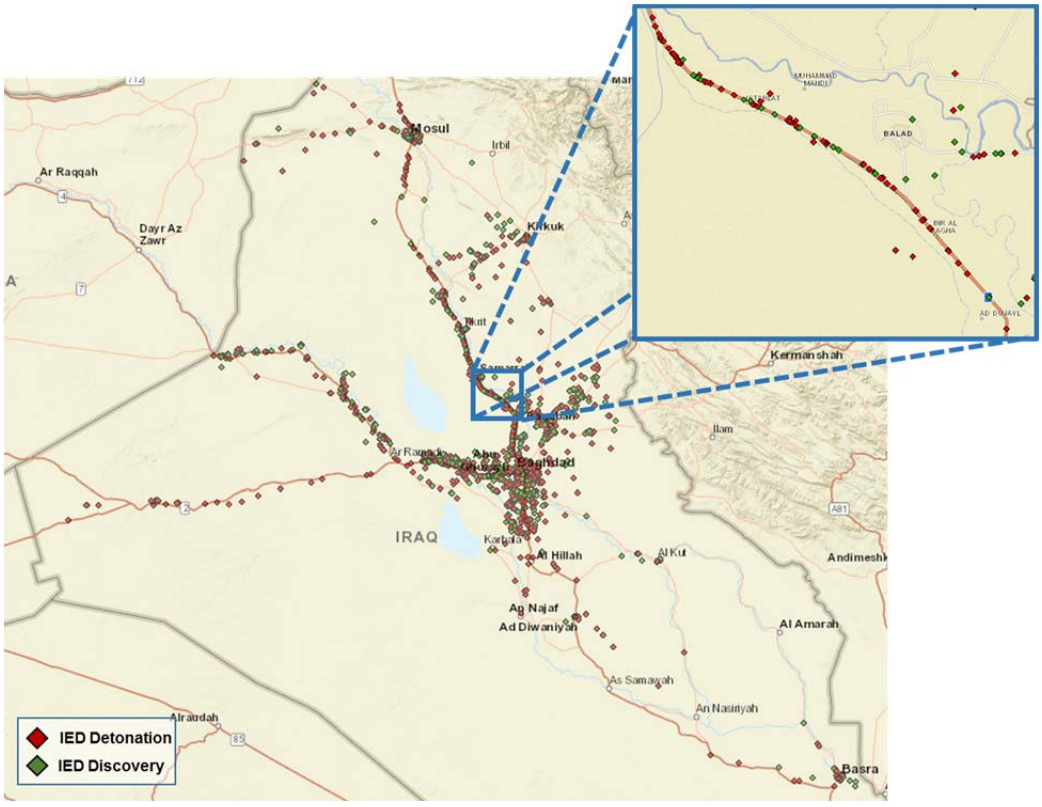
A. DATA

This thesis focuses on two data sources. The primary data source consists of all real-world IED events in Iraq from January, 2005 to December 2008 and is used to explore model performance against a realistic IED environment. The second data source is a notional dataset built with known patterns, which is used for model development, and testing.

1. Iraq Data

The U.S. military in Iraq and Afghanistan made significant efforts to capture as much information as possible about IED attacks. These significant activities (SIGACTS) were shared among units through various interlinked databases. The most prominent and widely used database was the Combined Information Data Network Exchange (CIDNE), which is maintained to this day by Central Command (CENTCOM) (CIDNE 2016). IEDs were considered SIGACTs and as such, the information collected at the scene of each IED resides within the CIDNE database. CIDNE also maintains the basic information about each event such as the date, time, and location, which are the three necessary fields used in this thesis. A Freedom of Information Act (FOIA) request to CENTCOM resulted in the acquisition of a subset of Iraq IED data.

The CIDNE database obtained from CENTCOM consists of date, time, location, and whether the device was detonated against its target or discovered before successful detonation. The data consists of every IED in Iraq from January, 2005 through December, 2008. There are ~49,700 detonations and ~32,100 discoveries totaling nearly 82,000 IED events during this four-year time period. Figure 3 is a map of IED activity during the month of September 2006, which was the most active IED month ever recorded in Iraq with 2768 IED events. The spatial representation of activity in this month mirrors that of other months with heavy concentrations of IEDs in urban areas such as Baghdad in Central Iraq and Mosul in Northern Iraq. IED activity outside of urban clusters follow linear paths which are the routes most often trafficked by coalition forces.



Heavy concentrations of IEDs in urban areas such as Baghdad (Central) and Mosul (North) with the remainder of activity focused along major and minor supply routes. Insert used to highlight how IED activity is spatially related to road networks.

Figure 3. September 2006 IED Activity in Iraq.

IED counts over 2005–2008 also prove to be indicative of the strength of the insurgency. From early 2005 until mid-2007 there is an increasing trend in the number of IED events and a corresponding growth in insurgency strength leading up to the surge. At the end of 2007, a period known as the “Sunni Awakening” or “Sons of Iraq” resulted in a significant decrease in SIGACT activity, specifically in central and northern Iraq, which can clearly be seen in Figure 4 (Wilbanks et al. 2010). The effect of the Sunni Awakening is clearly evident when this data is analyzed. There were a total of 16,256 IED events in the last 14 months of this dataset and a total of 32,323 events in the preceding 14-month period, a drop of nearly 50 percent. This drop in IED activity presents challenges to modeling IED patterns as IED activity is not stationary over this four-year horizon.

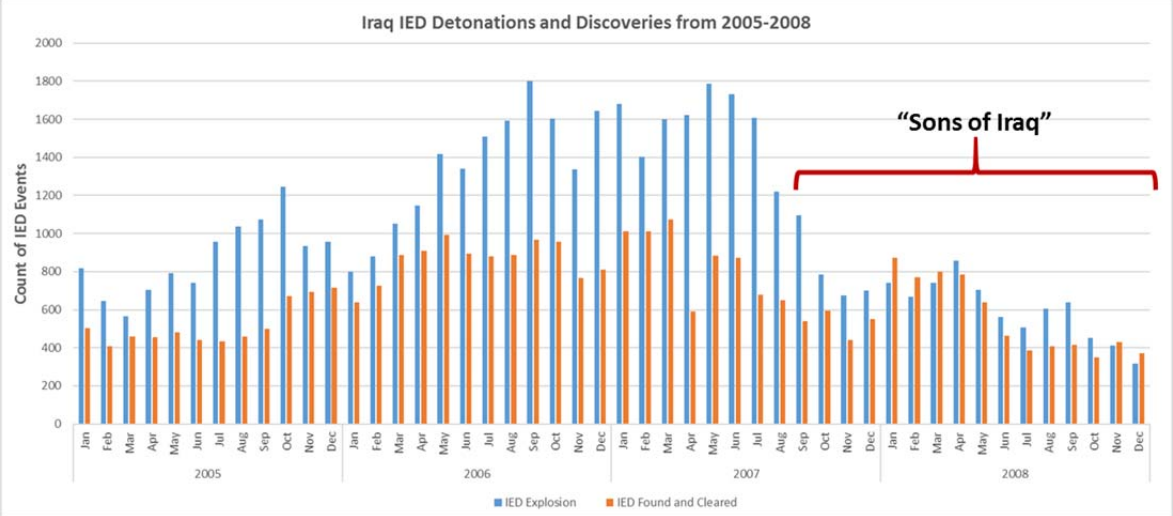


Figure 4. IED Detonations and Discoveries in Iraq from 2005–2008.

Lack of stationarity suggests that modeling IED activity as a homogenous Poisson process over this four-year period and across the extent of Iraq would not be appropriate. The mean number of IEDs per month across the entire dataset is ~1704 while the variance is a little over 354,000. If IED activity across the entire country and time period were well represented by a Poisson process, these two values would be nearly equal. This evidence against the homogenous

Poisson process is strengthened when one calculates inter-arrival time across the entire dataset. The mean inter-arrival time is 0.43 hours while the standard deviation is 0.697 across all 81,824 IED events. The fact that these two values differ by a reasonable margin for a fairly large dataset suggests that inter-arrivals do not come from an exponential distribution and therefore cannot be modeled with a Poisson Process.

Non-conformity to a Poisson process is further evidenced when the real-world dataset is filtered temporally and spatially, focusing on specific areas over shorter periods of time. For example, if we focus on the height of the insurgency between August 06 and July 07, IED activity looks relatively stationary (Figure 4), with a range of 2105 to 2768 IEDs per month. The mean number of IEDs per month was 2493 with a variance of 39,498. The inter-arrival mean was 0.292 with a standard deviation of 0.423. We also explored the possibility of IED activity supporting a Poisson process by spatially filtering the data around Baghdad proper during the height of the insurgency, and the analysis produced similar results. These results do not suggest that a Poisson process is a poor model for every possible spatio-temporal IED subset, but it does suggest that one should proceed with caution utilizing a Poisson process without furthering examining the specific data of interest.

2. Notional Data

The second dataset we explore is a notional IED database consisting of 106 IED events over a 31-day period. This dataset was created before work on this thesis began based on experiences in Iraq using known patterns reminiscent of those encountered when deployed. The IEDs in this dataset stretch across approximately 250 km of roadway with five named areas of interest (NAIs) drawn around high concentrations of IED activity (Figure 5). The spatial placement of NAIs in this notional environment is consistent with methods used in Iraq, which suggests that their placement is likely based on IED activity from over three

months ago. This reinforces the idea that spatially stagnant NAIs may not be the best method to spatially filter IED activity. The notional discovery versus detonation rate (percentage of IEDs discovered before successful detonation) is approximately 40 percent (Figure 6), which mirrors the rate seen in Iraq during the height of the insurgency from 2005–2007 as calculated from the real-world dataset described in the previous subsection.

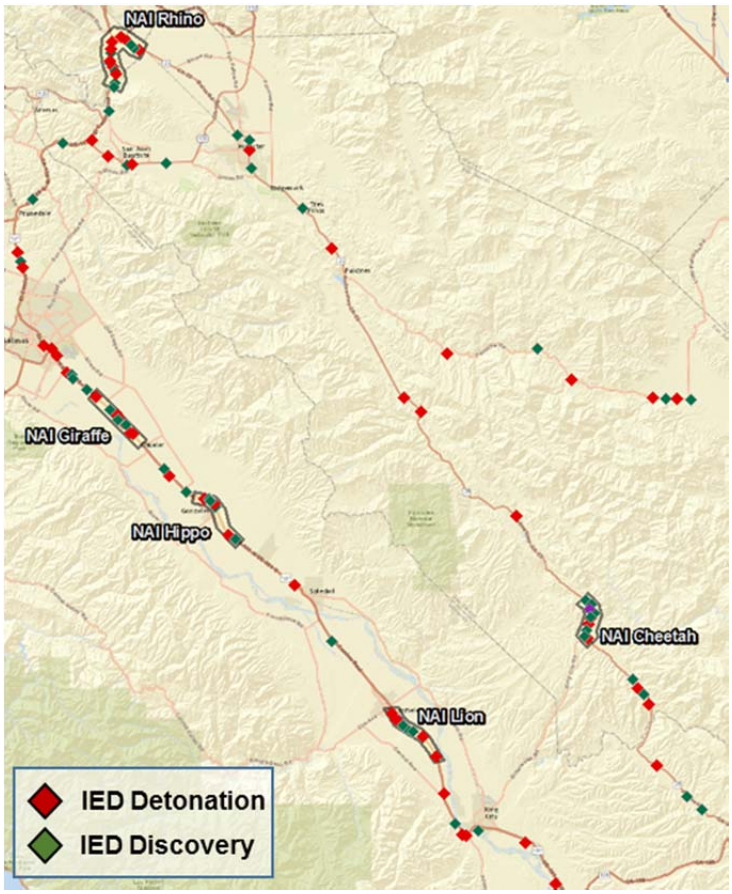


Figure 5. Spatial Representation of Notional Dataset.

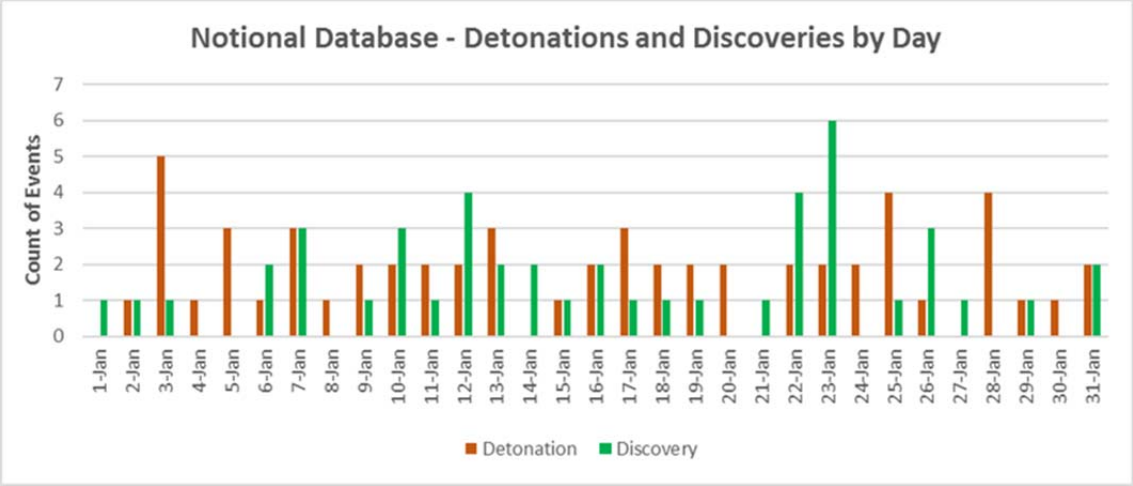


Figure 6. Detonations and Discoveries by Day from the Notional Dataset.

Table 1 and Figures 7–8 represent the data input and statistical output normally produced when analyzing IED activity for a single NAI. These were produced from the notional dataset and the primary inputs for traditional statistical analysis are time, location, and initiation type. Information about time and location provide the base upon which a route clearance schedule is developed, while initiation type allows a commander to allocate her counter-IED equipment according to the type of threat. This dataset is not based on real-world events. The notional dataset is primarily used to develop and test various prediction methodologies against known patterns while the real-world dataset is used to validate the chosen model.

Table 1. Notional IED Events in NAI Rhino.

MGRS	Date	Time	Date_Time	Type	MSR-ASR	NAI	Initiation
10SFF2904388465	3-Jan-16	14:19	1/3/16 14:19	Detonation	MSR Bulldog	Rhino	Command Wire
10SFF2891291170	3-Jan-16	19:24	1/3/16 19:24	Detonation	MSR Great Dane	Rhino	Unknown
10SFF2940987794	5-Jan-16	16:01	1/5/16 16:01	Detonation	MSR Great Dane	Rhino	Unknown
10SFF3096790780	11-Jan-16	23:01	1/11/16 23:01	Discovery	MSR Great Dane	Rhino	Unknown
10SFF2885489161	12-Jan-16	16:24	1/12/16 16:24	Detonation	MSR Great Dane	Rhino	Remote Controlled
10SFF2883590415	13-Jan-16	23:31	1/13/16 23:31	Detonation	MSR Great Dane	Rhino	Command Wire
10SFF2916386403	15-Jan-16	17:10	1/15/16 17:10	Discovery	MSR Bulldog	Rhino	Unknown
10SFF3132890475	22-Jan-16	5:06	1/22/16 5:06	Discovery	MSR Great Dane	Rhino	Remote Controlled
10SFF2876888916	22-Jan-16	13:42	1/22/16 13:42	Discovery	MSR Bulldog	Rhino	Unknown
10SFF2884289960	23-Jan-16	15:04	1/23/16 15:04	Discovery	MSR Great Dane	Rhino	Command Wire
10SFF3060491196	25-Jan-16	6:03	1/25/16 6:03	Detonation	MSR Great Dane	Rhino	Remote Controlled
10SFF2997391664	28-Jan-16	3:33	1/28/16 3:33	Detonation	MSR Great Dane	Rhino	Unknown
10SFF2931688081	30-Jan-16	15:17	1/30/16 15:17	Discovery	MSR Bulldog	Rhino	Unknown
10SFF3186490328	31-Jan-16	5:39	1/31/16 5:39	Detonation	MSR Great Dane	Rhino	Remote Controlled
10SFF2924786759	31-Jan-16	16:38	1/31/16 16:38	Discovery	MSR Bulldog	Rhino	Command Wire

IED database filtered to include NAI Rhino IED events. Database includes the type of event (detonation vs. discovery), the main supply route (MSR)/alternate supply route (ASR) on which the event took place, and the type of initiation method used by insurgents.

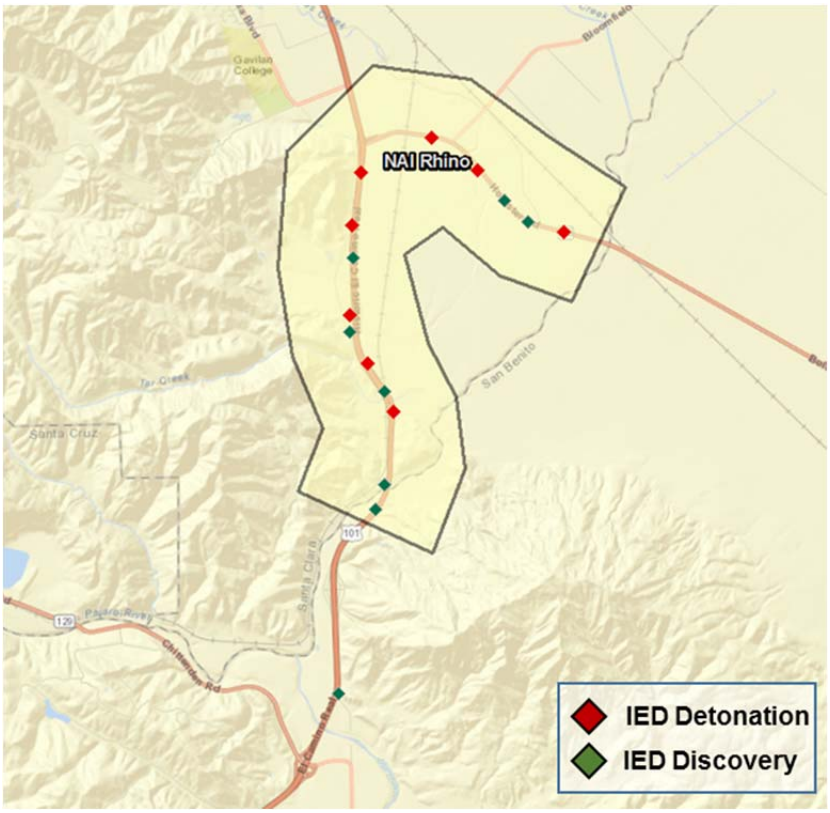
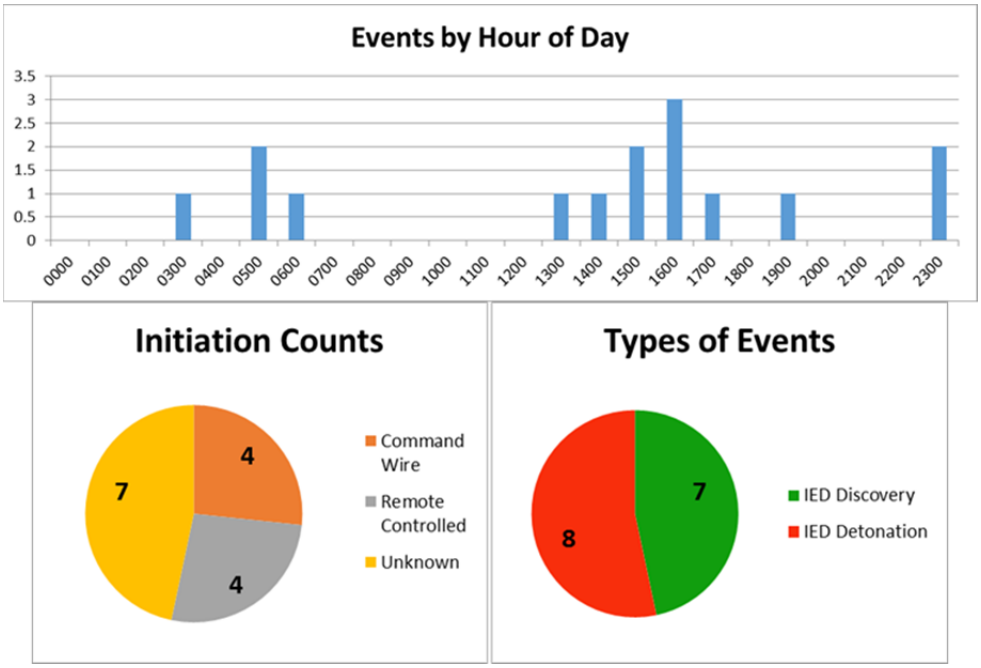


Figure 7. Spatial Representation of IED Events in NAI Rhino.



Time analysis suggests the existence of two primary attack times (0300–0600 and 1300–1700) while initiation type suggest two primary initiation types. This bimodal behavior should prompt a military intelligence analyst to break down events by time of day and initiation type to see if a spatial, temporal, and/or initiation relationship exists.

Figure 8. Traditional Statistical Output of IED Events in NAI Rhino.

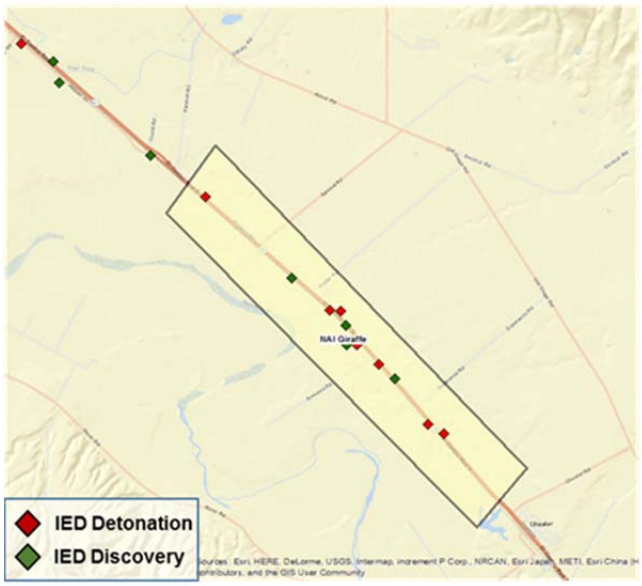
B. PATTERNS

Experience suggests that predictable IED patterns are most often present at a local level, both spatially and temporally. A simple example of an IED pattern is the appearance of an IED every other night for two weeks over a 5km stretch of road. It is very unlikely that patterns exist across more than 25 km of a road network or exceed two months of time. This spatial limit is based on insurgent coordination resulting in the physical separation of IED cells in rural areas as well an IED emplacer's potential unwillingness to travel long distances to conduct attacks. The temporal limit is based on the increased probability of pattern disruption as time passes. Potential IED pattern disruptions include the removal of coalition force presence or the death or capture of an IED cell member. These limitations require us to analyze sparse data (usually no more than 25 events) for pattern recognition. There is a greater chance to identify IED patterns and,

therefore, the next IED event, by relying only on the most recent activity in a given area.

1. Pattern Visualization

Deployed analysts developed a particular pattern visualization technique that allowed quick identification of the existence of a pattern given a sequence of IEDs. As an example of this technique, we construct Figure 9 from the notional dataset by filtering the data spatially to one particular NAI and temporally to the current month; this filtering results in a total of 11 events. Once we filter the data, the events are ordered sequentially and we compute the inter-arrival times as shown in the last column of Figure 9. We then plot the inter-arrival times using a scatter plot, with a smoother for quick visualization (Figure 10). The x axis represents the sequence index (1,2,3, . . .) and the y axis is the calculated inter-arrival time. Common patterns include a relatively flat line or cyclical oscillations.



Sequence	Date	Time	Time Between Events (in Hours)
1	3-Jan-16	3:22	22
	4-Jan-16	0:57	
2			29
	5-Jan-16	5:50	
3			17
	5-Jan-16	23:02	
4			131
	11-Jan-16	10:19	
5			22
	12-Jan-16	8:09	
6			21
	13-Jan-16	4:50	
7			147
	19-Jan-16	7:50	
8			172
	26-Jan-16	11:31	
9			11
	26-Jan-16	22:24	
10			29
	28-Jan-16	3:15	

Figure 9. Spatio-Temporal Subset with Inter-Arrival Calculations.

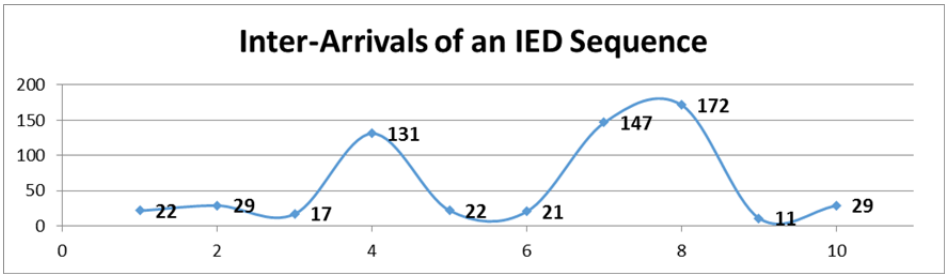


Figure 10. Time between Events Visualization Example.

2. Common Patterns

Identifiable IED patterns exist in two broad forms often based upon insurgent availability of funding, materials, and labor necessary for IED production. The first is common in areas with a near endless supply of IEDs and is the easiest to recognize. It is categorized by approximately constant inter-arrival times with low variability. Examples of this type of pattern are an IED every 22 to 26 hours for five days in a row or an IED every eight to 10 hours targeting a constant flow of logistical patrols. It is rare to see these patterns continue past six IEDs. Just as intelligence analysts are trying to identify patterns, insurgent elements want to avoid falling into easily identifiable patterns, and hence will often vary their emplacement times and locations. Using the visualization technique described in the previous section, this type of pattern would appear as a near flat line and prediction of the next event would be calculated using the mean inter-arrival time (Figure 11).

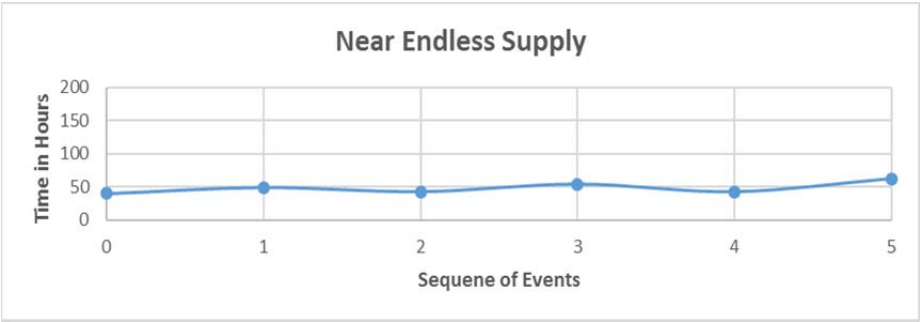


Figure 11. Near Endless Supply Pattern Visualization

The second and more common pattern is heavily influenced by supply and coalition force operations, and is therefore more cyclic in nature. These types of patterns may be less obvious to the insurgents conducting the attacks. This results in patterns involving many more IEDs than the patterns in the near endless supply scenarios previously described. However, these types of patterns are also more difficult for analysts to detect. Cyclical patterns can be further broken down into three categories: large supply, short supply, and steady supply. The exact form of the patterns differs significantly. Patterns emerging in large supply situations have multiple events in quick succession followed by a short lull in activity during resupply (Figure 12). Large supply cycles often occur spatially in IED hotspots. It is possible that these large supply cycles are the result of one IED supplier providing product to multiple individuals emplacing IEDs. The short supply scenario is the opposite with large amounts of time between most IED activity until a larger than normal supply becomes available (Figure 13). These patterns often appear in areas with lower levels of IED activity and are most likely the result of only one IED emplacer. Steady supply is the most predictable of these three cyclic patterns and most resembles a standard sinusoidal curve. These patterns typically have one to two events in short succession followed by a lull in activity as insurgents spread out the emplacement of their remaining IEDs while waiting for resupply (Figure 14). This is also the most common of the described predictable patterns and these patterns appear in both high and low density areas.

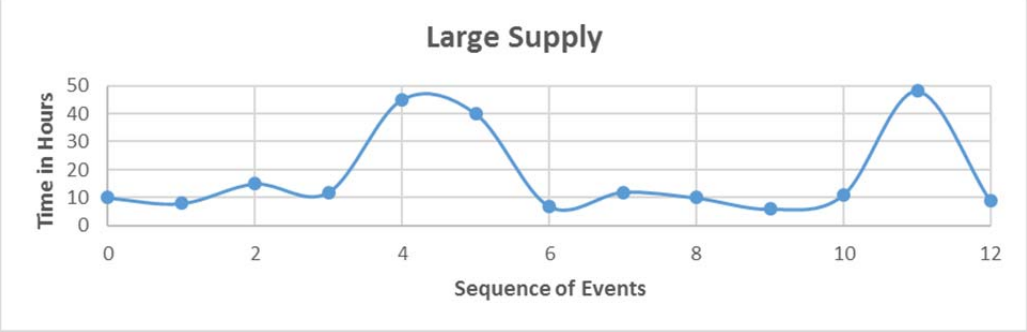


Figure 12. Large Supply Pattern Visualization.

Figure 13. Short Supply Pattern Visualization.

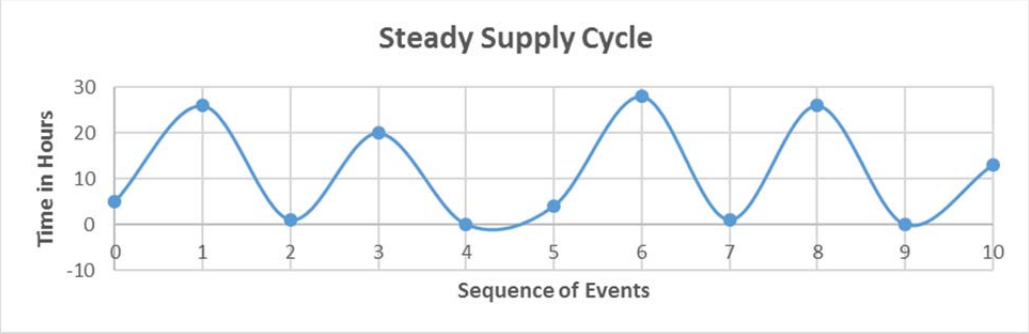
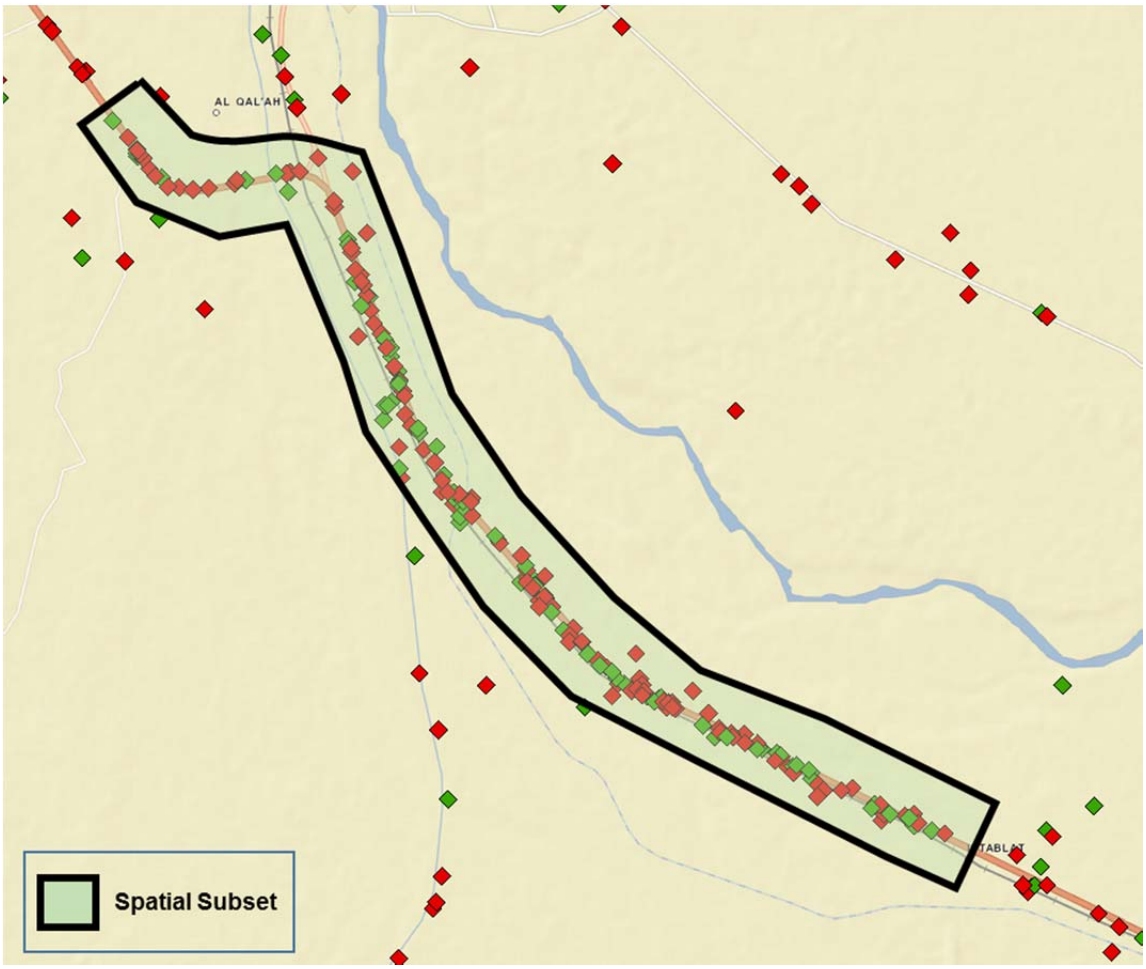


Figure 14. Steady Supply Pattern Visualization.

C. DATA FILTERING

When attempting to identify a specific pattern or signature based on time between IED events, it is important to spatially filter the data to a local level. NAIs would often be the basis for filtering even though they were rarely altered or reviewed. Other methods include using a bounding box described by an upper left and lower right geographic reference or hotspot analysis where the analyst begins their search in areas with the highest concentration of IEDs and gradually moves outward from the center of those clusters. This thesis will focus on the identification of temporal patterns once an analyst has already filtered the data spatially, leaving the criteria of the spatial filter up to the intelligence analyst (Figure 15*)*. This allows the analyst to focus on spatial pattern identification efforts in the broader context of personal knowledge and experience with the surrounding environment. It is the responsibility of the intelligence analyst to

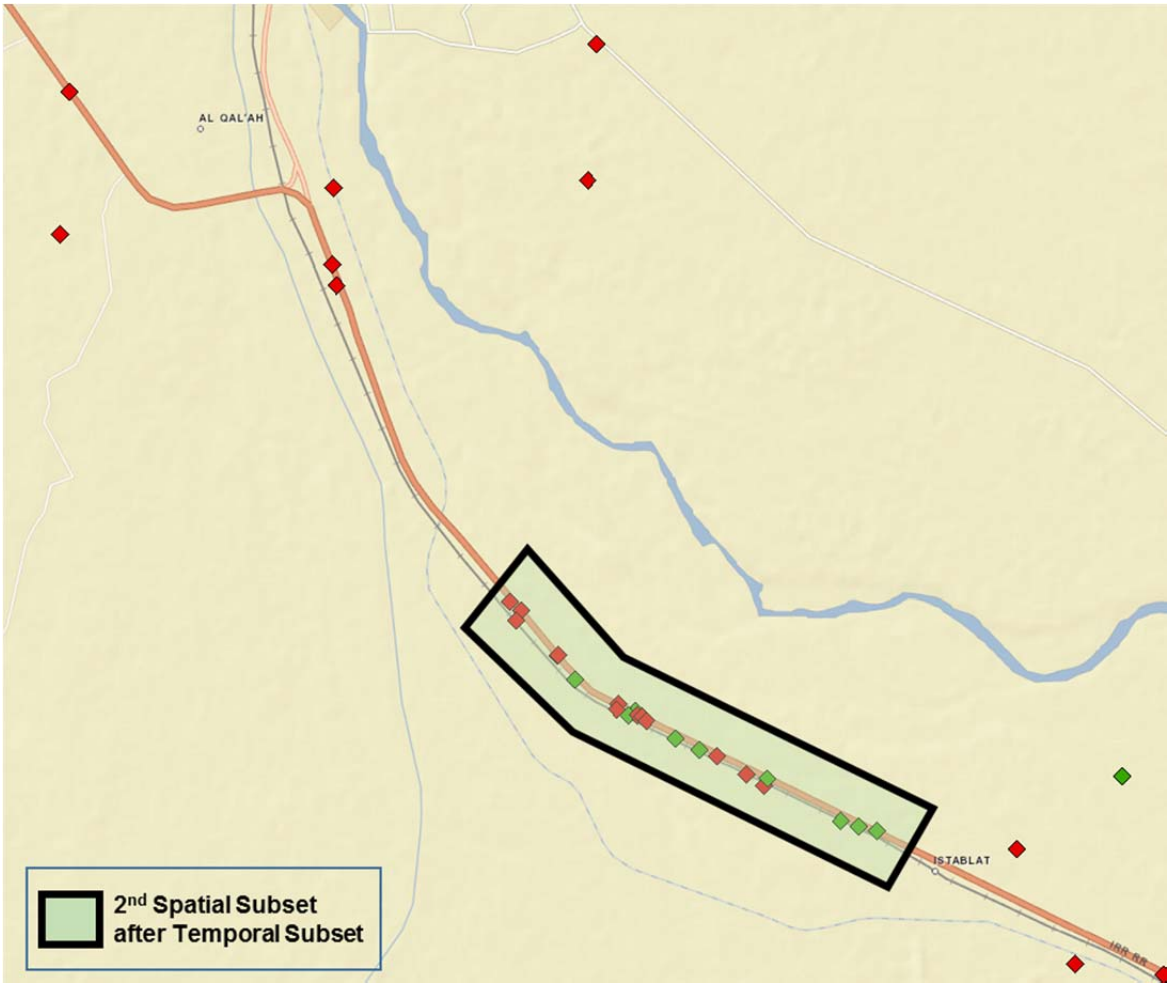
coalesce unquantifiable data into predictions of future enemy activity. An example of unquantifiable data in this context may be a key leader discussion with a local tribal leader who promises to address those in his tribe conducting attacks against U.S. forces. An engagement of this kind has the potential to disrupt or alter current IED activity in a given area. Allowing analysts to determine the spatial boundaries of their analysis gives them the necessary flexibility to analyze IED activity that is of concern to them and their leadership.



IED activity is filtered spatially as determined by the analyst. In this case, the analyst is concerned with IED activity along this particular stretch of HWY 1 due to the high casualty rate and they believe a pattern may exist since there is no reporting to suggest a disruption of IED activity in this area. This is a spatial representation of 2005 data over a 7km stretch of road (IED activity taken from Iraq dataset)

Figure 15. IED Spatial Filtering

Once the analyst filters the IED data spatially, it needs to be filtered temporally. Traditional military IED analysis tends to do this on a weekly or monthly basis since the counter-IED working groups met on a weekly or monthly basis. This approach is limited as it simply discards data based on date regardless of localized trends and patterns. It is likely to miss a slow developing pattern or overlook fast developing patterns because data selection is based on specific dates rather than localized trends. A better method would focus on the number of IED events for the spatial area previously identified by the analyst. Tactical IED pattern analysis is best performed when there are between six and 25 events. A discernable pattern lasting more than 25 events is unlikely, because the process underlying the pattern is likely to be disrupted, while anything less than six events is not enough data to make reasonable predictions. We recommend analysts should start with the last event in a localized area and backtrack to include no more than 25 events (Figure 16*)*. This thesis will focus on patterns defined by these quantity boundaries.



It is assumed for this example that the analysis is taking place on December 31, 2005. The analyst has already spatially filtered the area according to historic activity and now temporally filters the area by including the last 25 IED events, which reduces the number of IED events to be analyze. At this point, the analyst may decide to do another spatial filter if, for example, they notice that recent activity is focused in the southern portion of their original spatial filter. The data is now spatially and temporally filtered for pattern prediction.

Figure 16. IED Temporal Filtering.

The spatio-temporal filtering requires that we evaluate patterns at a very local level over a relatively short amount of time. Localized analysis is crucial to the prediction of specific locations and times of future events. Once data evaluation begins to exceed 40km of roadway or three months of data, an analyst is no longer attempting identify specific patterns but rather evaluating long-term risk along his assigned routes. Increasing the scope even wider to

activity over a year across a very large space then starts to feed the operational picture in terms of violence levels and whether coalition efforts are having the desired effect across the country. It is important to note that this thesis is limited to a very local level with the assumption that the individual unit hunting, or attempting to avoid the next IED, has very little ability to directly influence the operational campaign.

D. EXAMINATION OF FILTERED DATA

The results of the aforementioned spatio-temporal filtering will be an IED subset consisting of the most recent IED events (not to exceed 25) for a given area. If an IED pattern exists within that subset, it is likely that it will not be comprised of all events. A given subset of IED events will usually contain some "noise" events. In a deployed situation, IED noise may be generated by a newly arrived insurgent cell emplacing IEDs, a lack of coalition force presence for an unusual amount of time, or by an uncharacteristic IED supply change. It is also possible for multiple IED cells to operate in the same area without coordination. This occurs most often in urban environments. While the activity driven by each cell may produce a viable pattern, the combined activity across all cells may not. For our purposes, we define IED noise as IED events that fall within the temporal-spatial boundaries of interest, but do not contribute to a distinguishable pattern. These types of events may represent a temporary deviation from the traditional pattern in the area but do not necessarily suggest a complete disruption of that same pattern. It is then important to account for the possibility of IED noise in the final filtered dataset when developing pattern models.

The most straightforward method to determine whether a pattern exists within a larger collection of IED events is to iterate through every possible subset of the filtered IED data and evaluate whether each subset constitutes a pattern. This process, however, requires significant computing power since a filtered dataset containing between six and 25 IEDs results in over 33 million possible IED combinations. Depending upon the computational effort required to fit a

model, it may be infeasible to consider such a large number of subsets. Furthermore, differentiating between predictable IED patterns and random or unassociated IED events among so many combinations becomes a significant hurdle because of false alarms. The sheer number of subsets ensures that a model will likely specify that multiple subsets constitute a legitimate pattern, when in fact no actual underlying pattern exists.

Evaluating millions of possible IED combinations is outside the scope of this thesis; we suggest two simple rules to reduce the number of subsets for consideration. First is the removal of the last two events from the filtered data and setting them aside for model testing. In a filtered dataset with 22 inter-arrivals, we remove and set aside observations 21 and 22 while observations one to 20 would be used to develop the model and estimate appropriate parameters (Table 2*)*. The resulting model can then be used to predict observations 21 and 22. Model performance can then be measured by the difference between the model's predictions and the reality of observations 21 and 22. Testing against the last two inter-arrivals is consistent with techniques used in a deployed environment. An intelligence analyst would never suggest a change in patrol schedule or request surveillance assets in a situation where a possible pattern was identified even though the last two IED events deviated significantly from model predictions. This is not to suggest that the analyst would instantly dismiss the potential pattern but either the pattern should have already been identified through previous analysis or the analyst would want to confirm the patterns continuance at a future date before allocating limited resources. The first rule alone reduces the maximum number of possible combinations to just over 8 million. This is a method that will be discussed further at the beginning of Chapter IV.

Table 2. Examining Filtered Data.

Date/Time	Location	Type	Model Testing
11/10/2005 11:23	38SLC99507060	IED Found and Cleared	
11/10/2005 21:40	38SMC0052470061	IED Found and Cleared	
11/13/2005 13:13	38SMC040682	IED Found and Cleared	
11/15/2005 12:50	38SMC024690	IED Explosion	
11/21/2005 14:07	38SMC004700	IED Found and Cleared	
11/24/2005 14:21	38SLC99207100	IED Explosion	Must use 15 of remaining 20 events to evaluate potential patterns recognizing that up to 7 events could be IED noise
11/24/2005 14:21	38SLC9920671012	IED Explosion	
11/25/2005 16:45	38SMC043681	IED Found and Cleared	
11/29/2005 17:05	38SMC027688	IED Explosion	
12/4/2005 22:21	38SMC012696	IED Found and Cleared	
12/5/2005 10:00	38SMC019693	IED Explosion	
12/5/2005 10:20	38SMC00237018	IED Explosion	
12/9/2005 10:59	38SMC00557000	IED Explosion	
12/9/2005 11:11	38SMC0062869974	IED Explosion	
12/10/2005 11:00	38SMC002701	IED Explosion	
12/10/2005 15:54	38SLC9859071751	IED Explosion	
12/10/2005 15:55	38SLC984719	IED Explosion	
12/20/2005 11:50	38SMC0275368922	IED Found and Cleared	
12/20/2005 18:55	38SMC0461768029	IED Found and Cleared	
12/25/2005 17:33	38SMC007699	IED Explosion	
12/28/2005 11:32	38SLC98507159	IED Explosion	Used for Model Testing
12/29/2005 12:45	38SMC016694	IED Found and Cleared	

Continuing from Figure 16, this is the data subset selected for pattern prediction. The last two events (21 and 22) will be used for model testing while model development must include at least 15 of the remaining 20 events to be considered viable.

The second rule simply suggests the inclusion of a certain percentage of the filtered data when attempting to find patterns. This prevents a situation where very few IEDs of the original dataset coincidently generate a pattern. Instead, it forces the inclusion of a majority of IED activity in a given area for pattern distinction. At least 70 percent of IED events should be included in a given spatial and temporal subset to constitute a viable pattern. This would require potential subsets of the original 22 inter-arrivals to include at least 15 of those observations during the modeling process (see Table 2 for an example). This requirement for pattern recognition is reasonable in an operational environment without additional explanatory variables such as initiation types or biometrics, which have the potential to drive pattern analysis. This second step further reduces the number of possible combinations from just over eight million to a very manageable one hundred forty-six thousand.

There are many considerations intelligence analysts must take into account when filtering data for analysis. The remainder of this thesis will assume the analyst has subset the data with these considerations in mind and will focus on the development and testing of mathematical models to predict future IED activity given the data identified by an analyst. The next chapter describes each model in depth and provides a performance summary.

THIS PAGE INTENTIONALLY LEFT BLANK

III. MODEL DEVELOPMENT AND TESTING

A. INTRODUCTION

This chapter will focus on the evaluation of various mathematical methodologies to accurately model cyclical IED patterns. It will start by presenting the data we use to evaluate each methodology and then transition into a discussion about the methods we will consider. We will examine the Hawkes point process, Non-linear (NL) optimization of a sine function, and discrete Fourier transforms (DFT). For each methodology, we will briefly describe why we chose it, provide a detailed explanation of the mathematical process, and summarize the results of testing against known patterns.

B. DATA FOR TESTING

We first generated a dataset to test our methodologies. Rather than mining our data for ideal test patterns, we generate data that produced cyclical patterns similar to those described in Chapter II. This ensures that we can test our methodologies against every type of pattern (steady supply, high supply, and low supply), and that each pattern will consist of the same number of IED events. We chose to generate 15 events (producing 14 inter-arrivals) per pattern as the baseline for testing. We chose this number to balance between too few events, which could potentially be explained by coincidence, and too many events, which would increase the likelihood of pattern disruption. Generating our own patterns also allows us to control the presence of IED noise. For testing purposes, we did not include IED noise because we want to compare the performance of methodologies against one another in an ideal environment.

We first generated the data representing the steady supply pattern. This pattern consists of a sequence of “long” inter-arrival times, followed by a sequence of “short” inter-arrival times, followed by a sequence of “long” inter-arrival times, etc. We generated the short inter-arrival times using a uniform distribution between 20 and 30 hours, and we generated the long inter-arrival

times using a uniform distribution between 115 and 125. Each sequence consists of either one or two events with equal probability. This process creates a pattern defined by the presence of one or two IED events in short succession (between 20 and 30 hours) and one or two events in long succession (between 115 and 125 hours). Table 3 and Figure 17 illustrate the test pattern created using the process as described. Table 3 contains both the inter-arrival time (column 3) and the time of event (column 2); we use an arbitrary start date and time of 1/1/07 0:01 to derive the time. The cyclical nature of the data is clearly seen through the resultant sinusoidal curve (Figure 17). The mean and standard deviation of the inter-arrival times appear in Table 1 as well.

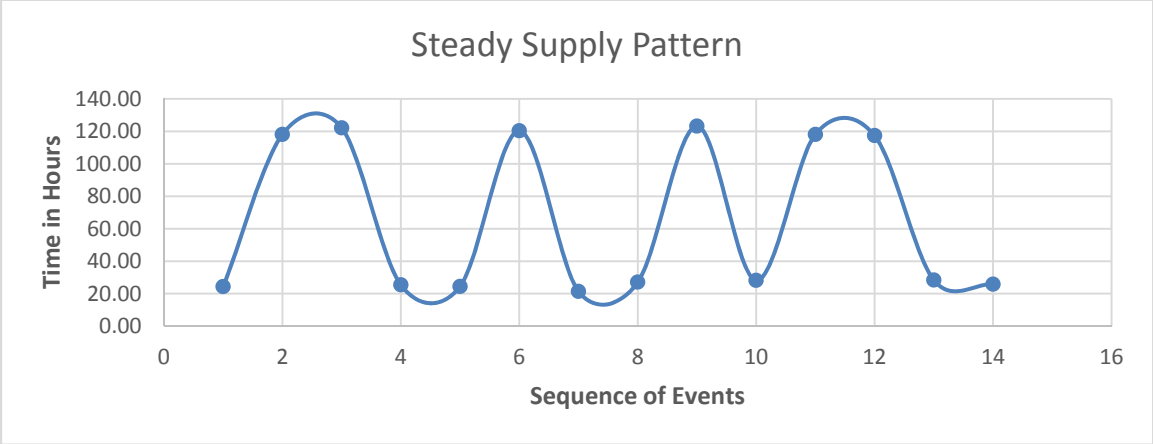


Figure 17. Visualization of Steady Supply Pattern.

Table 3. Data–Steady Supply Pattern.

Steady Supply Sequence		
Sequence	Date/Time	Time Between Events (hours)
	1/1/07 22:58	
1	1/2/07 23:17	24.32
2	1/7/07 21:23	118.11
3	1/12/07 23:32	122.15
4	1/14/07 0:54	25.37
5	1/15/07 1:19	24.41
6	1/20/07 1:38	120.31
7	1/20/07 23:02	21.41
8	1/22/07 2:09	27.11
9	1/27/07 5:23	123.24
10	1/28/07 9:31	28.13
11	2/2/07 7:37	118.10
12	2/7/07 4:59	117.35
13	2/8/07 9:18	28.32
14	2/9/07 11:06	25.80
	Mean	66.01
	STD Deviation	46.70

We use a similar process to generate the patterns for high supply and low supply scenarios (Figures 18–19). Both require the use of one additional uniform distribution to differentiate between how many events occurred during the long inter-arrival period versus the short inter-arrival period. For the high supply pattern, the long inter-arrival periods consist of either one or two events while the short inter-arrival period contained either four or five events. This translates into four or five events occurring in quick succession (between 20 and 30 hours) with a resupply cycle represented by one or two events with long inter-arrivals (between 115 and 125). The reverse holds true for the low supply pattern. The resultant data and curves appear in Tables 4–5 and Figures 18–19.

Table 4. Data–High Supply Pattern.

High Supply Sequence		
Sequence	Date/Time	Time Between Events (hours)
	1/2/07 20:59	
1	1/3/07 17:14	20.25
2	1/4/07 22:36	29.38
3	1/5/07 22:26	23.82
4	1/7/07 4:59	30.56
5	1/8/07 4:51	23.87
6	1/13/07 3:52	119.01
7	1/18/07 6:11	122.31
8	1/19/07 11:52	29.69
9	1/20/07 9:44	21.87
10	1/21/07 10:00	24.26
11	1/22/07 11:10	25.18
12	1/23/07 17:05	29.92
13	1/28/07 21:13	124.13
14	1/29/07 22:42	25.49
	Mean	46.41
	STD Deviation	39.50

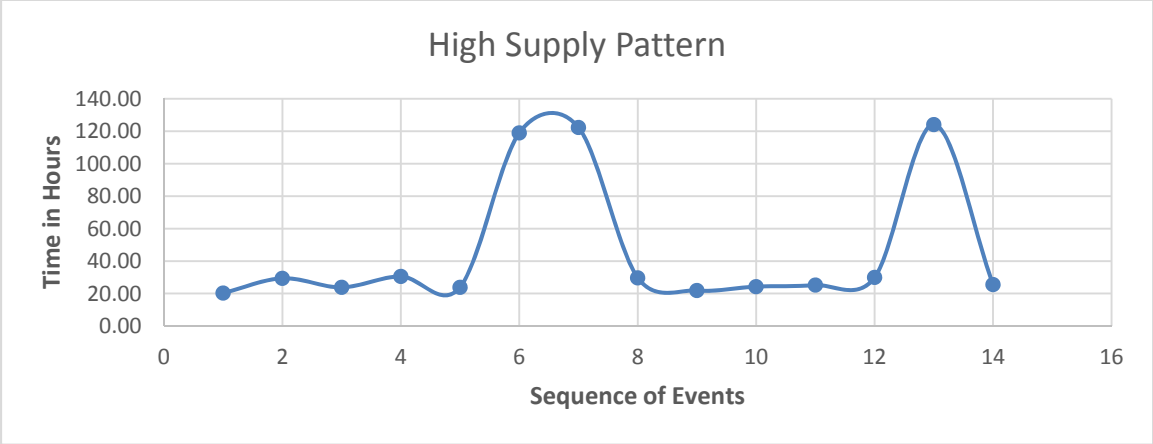


Figure 18. Visualization of High Supply Pattern.

Table 5. Data–Low Supply Pattern.

Low Supply Sequence		
Sequence	Date/Time	Time Between Events (hours)
	1/3/07 4:52	
1	1/8/07 9:37	124.76
2	1/13/07 6:27	116.84
3	1/18/07 10:19	123.87
4	1/23/07 10:33	120.24
5	1/24/07 11:32	24.98
6	1/29/07 8:49	117.29
7	2/3/07 13:45	124.93
8	2/8/07 12:08	118.37
9	2/13/07 8:12	116.06
10	2/14/07 9:09	24.96
11	2/15/07 7:55	22.76
12	2/20/07 8:13	120.30
13	2/25/07 8:08	119.91
14	3/2/07 4:30	116.37
	Mean	99.40
	STD Deviation	39.36

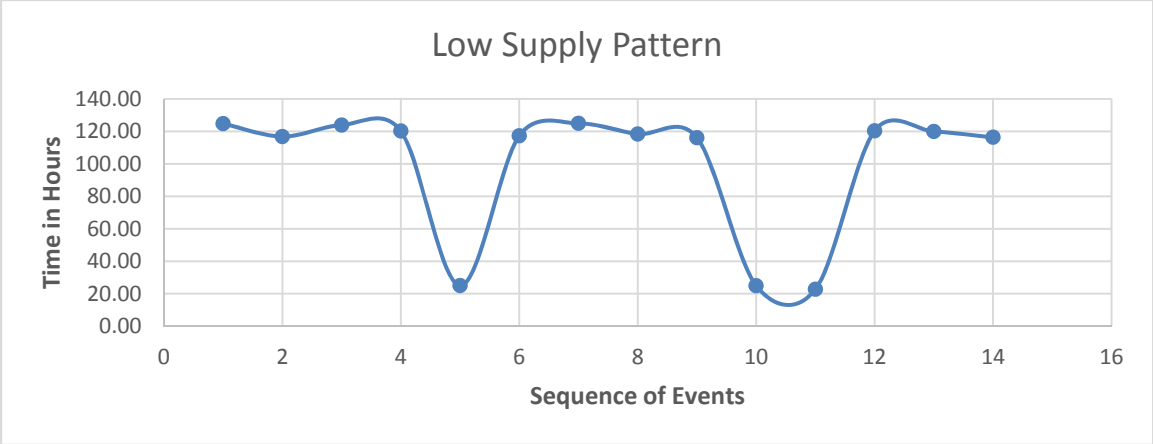


Figure 19. Visualization of Low Supply Pattern.

We also generated a random series of 15 IED events with inter-arrival times uniformly distributed between 0 and 100 hours (Table 6 and Figure 20*)*. We test the randomized series using the same methodologies to compare against the well-established patterns. The difference in model performance between the randomized data and known patterns is an indicator of how well a particular methodology distinguishes between patterns and non-patterns.

Table 6. Data–Random Sequence.

Random IED Sequence		
Sequence	Date/Time	Time Between Events (hours)
	1/7/07 2:48	
1	1/8/07 8:33	29.75
2	1/11/07 18:04	81.53
3	1/15/07 0:38	78.56
4	1/17/07 17:41	65.06
5	1/20/07 9:34	63.88
6	1/21/07 19:49	34.24
7	1/25/07 12:37	88.80
8	1/26/07 4:54	16.28
9	1/27/07 20:29	39.60
10	1/28/07 8:20	11.84
11	1/28/07 18:10	9.84
12	2/1/07 11:30	89.33
13	2/4/07 12:02	72.53
14	2/6/07 23:41	59.65
	Mean	52.92
	STD Deviation	27.69

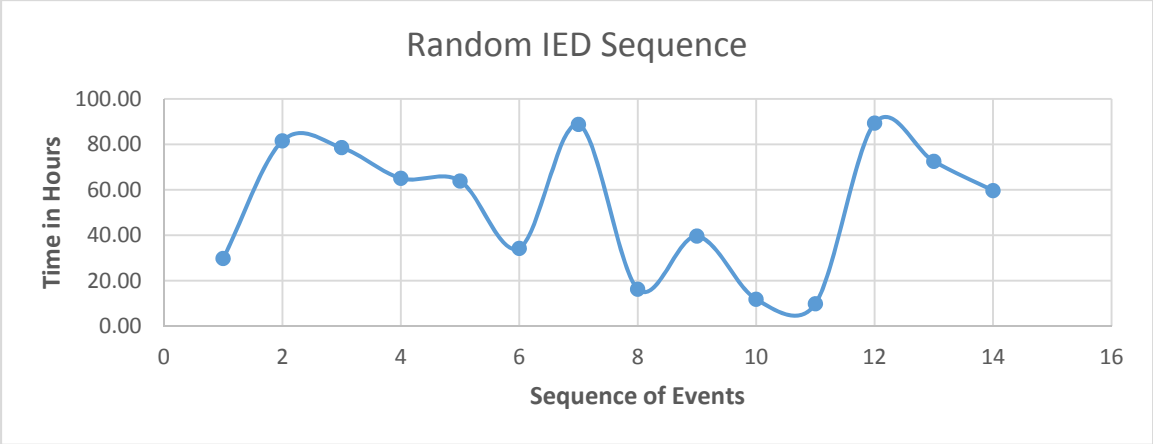


Figure 20. Visualization of Random Sequence.

C. MEASURING MODEL PERFORMANCE

We primarily measure model performance using root-mean-squared error (RMSE). The formulation of RMSE appears in Equation 1, where n represents the number of inter-arrivals into the system (number of IED events – 1),

y_i represents the recorded inter-arrival times in hours, and $\hat{y}_i$ represents the estimated inter-arrival provided by the model in question.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (1)$$

We will also plot the fitted values $\hat{y}_i$ on the same plot as the actual inter-arrival times y_i to visually inspect the model fit. RMSE provides a mathematical approach to compare models and IED sequences; however, it does not capture where along a sequence the model performs well or poorly. Visual inspection during development and testing allows us to identify whether a large RMSE (poor model fit) is the result of a model's inability to capture outlier values or the result of consistent errors even though the model captures the general shape of the pattern.

D. HAWKES POINT PROCESS

The first methodology we evaluate is the Hawkes point process. In the early 1970s, Alan Hawkes developed a theoretical model of a self-exciting point process (SEPP) (Hawkes 1971). A classic application of such a model is earthquakes triggering aftershocks (Ogata 2012). When applying SEPP to earthquake behavior, the likelihood of earthquakes occurring in the near future increases after an initial earthquake event. There have also been successful attempts to model crime patterns with Hawkes processes, where an initial crime triggers a flurry of additional crime (Mohler 2009). The same logic may apply to cyclical IED patterns. For example, in the high supply pattern scenario, an initial IED after a lull in activity is then followed by several more IEDs in quick succession. While the Hawkes dynamics may reasonably represent the high supply pattern, it may not do as well for other types of supply patterns.

A traditional point process consists of a single arrival rate (or intensity function) given by Equation 2, where N(t) represents the number of IED events by time t, and {N(t), t>0} is a counting process (Toke 2011).

$$\lambda(t) = \lim_{h \to 0} \frac{1}{h} P[N(t+h) - N(t) > 0 \mid N(s), 0 < s < t] \quad (2)$$

As such, λ(t) is the instantaneous arrival rate of new events given the history of events. Another interpretation is that the conditional probability an event will occur in the next small time period h is roughly λ(t)*h (Toke 2011). A homogenous Poisson process has λ(t) = λ for all t and histories.

Hawkes point process differs from a Poisson process in that an arrival at time s, increases the intensity function λ(t) for t > s. Each new arrival into a Hawkes point process triggers an increase in the intensity function, creating a situation where the current arrival rate is determined by previous activity (Hawkes 1971). The specific intensity function we will consider appears in Equation 3 (Toke 2011).

$$\lambda(t) = \mu_0 + \sum_{t_i < t} \alpha e^{-\beta(t - t_i)} \quad (3)$$

The parameter μ_0 represents the baseline intensity; IEDs will arrive at rate μ_0 if no other IEDs events have occurred for an extended period of time. Roughly speaking, if it has been a long time since the last IED event, then the time until the next IED event has an exponential distribution with rate μ_0. The summation term in Equation 3 allows for an increase in the intensity function whenever a new IED event occurs. However, the impact of an IED event on the arrival rate of future IED events diminishes over time. The β parameter captures the decay rate of the arrival rate change, which is a measure of the duration of influence a new arrival has on the intensity function (larger β implies the influence is fleeting). The α term is a measurement of how influential a new arrival is to the system. Large α implies a new arrival will have an immediate and significant impact on the intensity function, which may trigger a cascade of future events. An α close to zero suggests that the process could be modeled adequately as a Poisson process (Toke 2011).

1. Description of the Methodology

Given a set of data, we can estimate the three Hawkes parameters. To do so, we will take a maximum likelihood estimation (MLE) approach. We adapted the log-likelihood function from (Toke 2011), which appears in Equations 4–5.

$$L = -t_n\mu + \frac{\alpha}{\beta}\sum_{i=1}^{n}(e^{-\beta(t_n - t_i)} - 1) + \sum_{i=1}^{n}\ln(\mu + \alpha R_i) \tag{4}$$

$$R_i = (1 + R_{i-1})e^{-\beta(t_i - t_{i-1})} \tag{5}$$

Equation 5 represents an intermediate value necessary to solve the negative log-likelihood function (Equation 4). To estimate the three parameters μ_0, α and β requires formulating a non-linear (NL) optimization problem to maximize the log-likelihood function.

We perform the MLE computations in the R programming language using the general optimization function from the **`stats`** package (Appendix A. Scripts). The inputs for the optimization function are initial parameter estimates for μ_0, α and β, the negative log-likelihood function as an R function, and a vector of recorded arrival times. It is important to note that we use arrival times instead of inter-arrival times in the calculation of the likelihood function for the Hawkes process. The optimization routine returns the "optimal" estimates of μ_0, α and β. As the optimization problem is non-linear the optimization routine may terminate before finding the global optimum. We re-run the optimization routine 1000 times with different initial estimates for μ_0, α and β. We generate initial parameters from a random uniform distribution over 0 to 10, and we only keep the iteration that produces the smallest value of the negative log-likelihood. The only remaining constraint is forcing a non-negative α and β.

After calculating estimates for μ_0, α and β, we can evaluate the model performance by computing the RMSE. To do this, we perform a modified calculation of Equation 1. We first simulate a Hawkes process for the given values of μ_0, α and β, which produces our estimated inter-arrival times $\hat{y}_{ij}$ where i represents the event in sequence and j represents an individual simulation run.

We simulate a Hawkes process using the **hawkes** R package. We then compute RMSE using Equation 1 for this one simulated process after reconverting arrival times into inter-arrival times. We repeat this process 1000 times, computing 1000 RMSEs, taking the average to arrive at our final measure of effectiveness for this method (Equation 6).

$$RMSE = \frac{1}{1000}\sum_{j=1}^{1000}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_{ij})^2} \tag{6}$$

2. Model Performance

The resulting RMSE from our testing appears in Table 7. We also plot the mean arrival time for each event (computed via simulation) against the actual data in Figure 21. Visual inspection shows that the Hawkes model fit defaults to a single mean arrival rate for the system. The Hawkes optimization assigns an optimal value of zero to α for all scenarios, which suggests that a traditional Poisson process fits better than a more complex Hawkes process. Just using the mean arrival rate will not capture the cyclic nature of most IED patterns, as illustrated in Figure 21. Visualization of Hawkes performance against low supply, steady supply and the random data can be found in Appendix B. Visualization of Model Results.

Table 7. RMSE Results and Parameters from Modeling IED Behavior Using Hawkes Point Process.

				RMSE
Pattern	μ	α	β	All Data
Random	0.0189	0	43.42	56.69
Steady Supply	0.0151	0	43.16	79.2
High Supply	0.0215	0	94.69	58.51
Low Supply	0.01	0	82.2	103.29

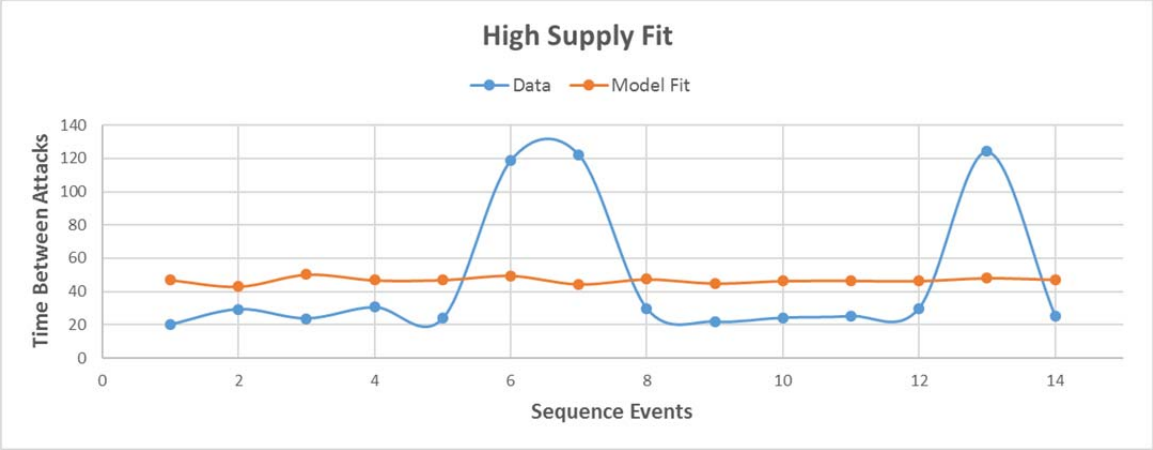


Figure 21. Model Fit by Hawkes Point Process of the High Supply Pattern.

Our findings suggest that a Hawkes point process is not a viable method to model IED patterns. It is likely that Hawkes is unable to accurately model existing IED patterns due to the small number of observations; most work that studies self-exciting point processes considers at a minimum thousands of events (Masuda et al. 2012, Fox et al. 2015, Lewis et al. 2011). Hawkes point processes are commonly used with financial data, which capture observations every second, or millisecond of a trading day (Toke 2011). The result is a dataset with thousands to millions of data points that gradually fluctuate over time. As an example, Hawkes point process added predictive power when modeling civilian casualties in Iraq over the extent of the war but the dataset consisted of nearly sixteen thousand observations (Lewis et al. 2011). This is a stark contrast to the six to 25 IED events that we focus on here.

E. NON-LINEAR (NL) OPTIMIZATION OF SINE CURVES

Visual inspection of known patterns, specifically the pattern established in steady supply situations, reveals a sinusoidal-type curve (Figure 17). We next focus on fitting a curve using a sine function with the hope that the model would not only fit known patterns well, but be able to differentiate known patterns from

random data. Two techniques emerged during our research and both are based upon the same base equation (Equation 7) (Dunbar 2005).

$$y_i = A\sin(f(x_i - \phi)) + K \tag{7}$$

Both require the estimation of four parameters: Amplitude (A), which is a measurement of the range of time between events; Frequency (*f*), representing how quickly the sine curve repeats; Phase-shift (φ), which captures where in the sin period the sequence starts; and Offset (K), which should correspond approximately to the mean or median of the inter-arrival time distribution.

The first technique involves a NL optimization that solves for all necessary parameters; however, it is computationally expensive with long run times. The second technique significantly reduces run time by estimating K (offset) as the inter-arrival mean (Equation 8), and A (amplitude) as half the range of the inter-arrival data (Equation 9).

$$K = \frac{\sum_{i=1}^{n} y_i}{N} \tag{8}$$

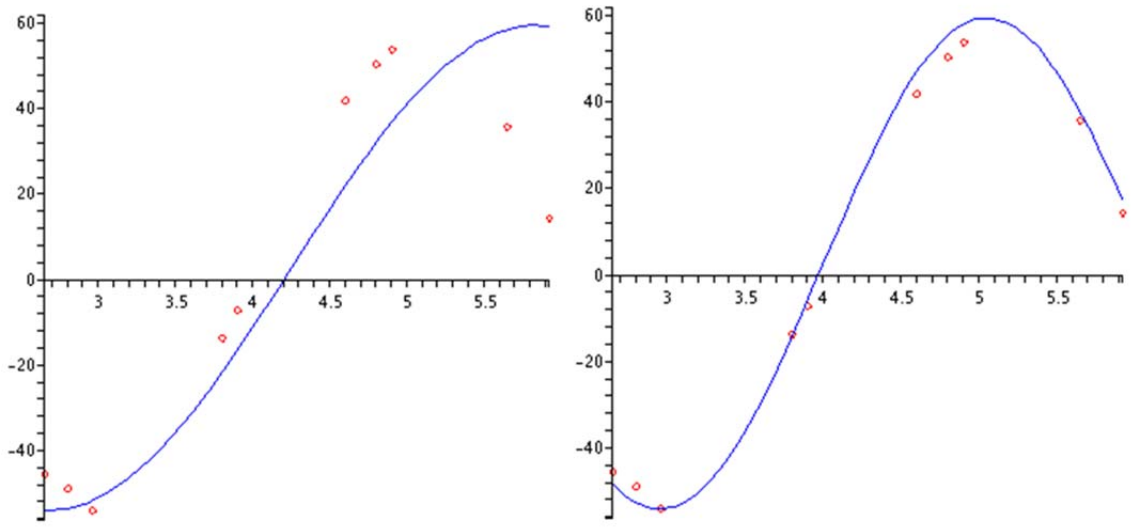
$$A = \frac{(\max y_i - \min y_i)}{2} \tag{9}$$

Estimation of the offset and amplitude allow us to transform the original equation into a linear regression (Equation 10), since we now have estimates of A and K, and y_i is the recorded inter-arrival times. The right side of the Equation 10 has the phase-shift and offset as the only unknown parameters remaining.

$$\arcsin\frac{y_i - K}{A} = (x_i + \phi) f \tag{10}$$

Computational run times for a linear regression are trivial in comparison to a NL optimization but the NL optimization method generally produces a more accurate model fit when working with the small datasets we consider (Figure 22) (Dunbar 2005). After seeing the poor fit using the Hawkes point process methodology, we pursue the option that provides the best model fit, choosing the

NL optimization technique. This leaves the option of exploring the linear regression technique if run times prove to hinder our ability to fully analyze an area.



Left Graph—Fit using estimation and transformation. Right Graph—Fit using NL optimization.

Figure 22. Sine Curve Fitting. Source: Dunbar (2005).

1. Description of the Methodology

Non-linear optimization (also known as non-linear programming) is a process that attempts to maximize or minimize an objective function by manipulating real variables over sets of constraints. Linear programs (LPs) and NLPs are normally calculated using a variety of algorithms that constitute a model solver built into various software packages. We choose to use the solver provided by Microsoft Excel as it is readily available to deployed military analysts, unlike many other optimization solvers that require expensive licenses. Moreover, its operation does not require the analyst to learn or understand a complex coding language.

This particular NLP requires the use of only one index, which is the sequence of IED events, and one set of given data, which is the recorded time between IED events in hours.

Index Use

i ϵ I Sequence of IED events

Given Data

y_i recorded time between each IED event (hours)

The decision variables associated with this problem are the amplitude, frequency, phase-shift, and offset. The NLP algorithm will manipulate these variables to find an optimal solution to the objective function. Similar to the Hawkes Point Process, using the NLP solver in Excel requires initial estimates of the parameter values. We perturb these initial estimates of the decision variables for every iteration to increase our chances of finding the global minimum. We initialize the solver 50 times for each pattern test with random uniform starting values for the decision variables to mitigate the risk of constantly reporting the same local minimum.

Decision Variables

Amplitude (A)	measurement of the range of time between events [hours]
Frequency (*f*)	measurement of how quickly the sine curve repeats [radians/sequence]
Phase-shift (φ)	measurement of how far sine curve should shift along the sequence axis [sequence]
Offset (K)	measurement of how far sine curve should shift along the time between event axis [hours]

The final formulation consists of the objective function (Equation 11) and the constraints placed on the decision variables. Although this is an unconstrained problem, the solver built into Excel has difficulty solving this NL optimization without constraining the decision variables, which led us to develop

a set of very broad but logical constraints (Equations 12–15). We know that K is roughly the inter-arrival mean and A is roughly half the inter-arrival range so constraining those values from zero to 200 and -100 to 100, respectively, does not hinder the optimization. The phase-shift is constrained to the maximum number of IED events we will analyze. The last constraint simply ensures the model does not produce negative inter-arrival times (Equation 16). Lastly, the objective function simply attempts to minimize the RMSE (Equation 11).

Formulation

$$\underset{RMSE}{MIN}\sqrt{\frac{1}{n}\sum_{i=1}^{n}[y_i - A\sin(f(i-\phi)) + K]^2} \tag{11}$$

$$-100 \le A \le 100 \tag{12}$$

$$-25 \le f \le 25 \tag{13}$$

$$-25 \le \phi \le 25 \tag{14}$$

$$0 \le K \le 200 \tag{15}$$

$$A\sin(f(i-\phi)) + K \ge 0 \tag{16}$$

2. Model Performance

NL optimization of sine waves performed significantly better than Hawkes point process against all patterns we tested, including the random data. It captures the general curvature of all patterns and results in lower RMSE for all IED combinations (Table 8). The model fit of the steady supply pattern provides a good example of these results (Figure 23) and the remaining result visualizations appear in Appendix B. Visualization of Model Results. However, the NL approach has several disadvantages.

Table 8. Optimized Parameters and RMSE Results from Modeling IED Behavior Using NL Optimization of Sine Waves.

Pattern	Amplitude (A)	Frequency (f)	PhaseShift (ø)	Offset (K)	RMSE All Data
Random	25.183	18.247	-18.121	49.978	21.714
Steady Supply	-55.645	-14.620	-15.628	70.152	26.252
High Supply	38.242	-19.768	18.172	45.998	28.993
Low Supply	-44.288	17.706	2.077	93.745	24.930

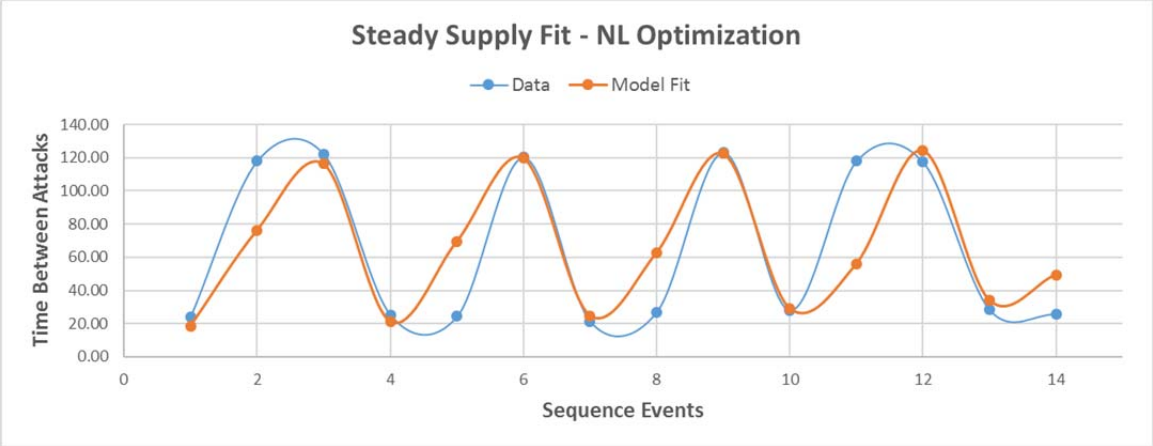


Figure 23. NL Optimization Model Fit of Steady Supply Pattern

One disadvantage, as previously mentioned, is the long computational run times associate with NL optimization. We were regularly running into solver run times of over a minute for a single test (without perturbation) during the testing process. This presents a substantial issue when, as described in Chapter II, it may be necessary to test over one hundred forty-six thousand possible IED combinations in a sequence containing 25 events. There are three possible methods to reduce run-times using the built-in Excel solver. The first method involves lowering the constraint precision, which prevents the solver from re-evaluating the objective function after miniscule changes in the decision variables. Another method is to change the convergence value, which ends a solver iteration if the solver cannot improve the objective value by the given threshold. Lastly, the option exists to place a simple time limit on how long a

single iteration can run. We used all these methods to significantly reduce our run times, changing the constraint precision to 0.001, the convergence value to 0.01 and placing a limit of one second per iteration. However, using NL optimization to evaluate one hundred forty-six thousand possible combinations with only 10 perturbations would take over 400 hours even with these constraints in place.

The other disadvantage is the model's inability to distinguish between random data and known patterns. The resulting RMSEs (Table 8) suggest that the model fits the random data best of the four patterns considered. As discussed in Chapter II, one possible technique to help distinguish between pattern and IED noise is the establishment of the last two inter-arrival times as the test set which should better highlight the predictive abilities of the models. We explore this concept in Chapter IV.

F. DISCRETE FOURIER TRANSFORMS

The computational expense associated with the NL optimization prompted us to search for an additional methodology to efficiently model sinusoidal patterns. This led us to Discrete Fourier Transforms (DFTs). DFT is a particular algorithm associated with Fourier analysis that consists of decomposing a signal into individual harmonics (Smith 1997). It represents functions in terms of sines and cosines, which is appealing to IED application as we have a natural cyclical form to the most common types of IED patterns. Just as data is often fit to a polynomial function, which is a linear combination of power functions, the DFT represents the data as a linear combination of sine and cosine functions, which we can use to capture cyclical behavior. As previously mentioned in Chapter I.4, Helms described the technique of using DFTs to model crime activity and we believe it is an appropriate technique to model IED activity.

1. Description of the Methodology

The general formulation of a DFT function appears in Equation 17 (Smith 1997). It consists of a single constant, (N-1)/2 cosine terms, and (N-1)/2 sine terms where N represents the number of inter-arrivals in the IED combination.

$$\hat{t}_j = \frac{a_0}{2} + \sum_{h=1}^{(N-1)/2} a_h \cos\left(\frac{2\pi j}{N} h\right) + \sum_{h=1}^{(N-1)/2} b_h \sin\left(\frac{2\pi j}{N} h\right) \quad (17)$$

Similar to a polynomial regression, a DFT model requires the estimation of the a_h and b_h parameters. Fortunately, there is a closed form equation to solve for these parameters as highlighted by Equations 18–19 where t_j represents the recorded inter-arrival time between two events, h represents a single harmonic whose range is defined by Equation 20, and j represents the index of events where j=0 for the first inter-arrival.

$$a_h = \frac{2}{N} \sum_{j=0}^{N-1} t_j \cos\left(\frac{2\pi h}{N} j\right) \quad (18)$$

$$b_h = \frac{2}{N} \sum_{j=0}^{N-1} t_j \sin\left(\frac{2\pi h}{N} j\right) \quad (19)$$

$$h = 0,1,\ldots,\left\lfloor \frac{(N-1)}{2} \right\rfloor \quad (20)$$

To prevent over-fitting, we will limit our exploration to DFT models with three terms (constant, one sine term, and one cosine) and with five terms (constant, two sine terms, and two cosine terms). We will refer to a DFT model with three terms as a single harmonic model and the model with five terms as a dual harmonic model for the remainder of this thesis.

If we only want a model with one or two harmonics, we do not need to re-calculate new parameters for each model; we can still use Equations 18–19. We just need to determine which index to use for the single harmonic model and which two indices to use for the dual harmonic model. Equation 21 allows us to rank-order harmonics, for h>0, according to their ability to summarize the data.

The maximum value returned by Equation 21 highlights the most influential harmonic and is defined as H_1 in subsequent equations.

$$a_h^2 + b_h^2 \tag{21}$$

Once we determine the most influential harmonic we can substitute the values of a_H and b_H into Equation 22 to estimate the inter-arrival times using a single harmonic model. If we would like to use a dual harmonic model, Equation 23 adds the necessary terms where H_1 represents the most influential harmonic and H_2 represents the second most influential harmonic. Each additional harmonic beyond the single most dominant harmonic adds two terms to the model.

$$\hat{t}_j = \frac{a_0}{2} + a_{H_1} \cos\left(\frac{2\pi j}{N} H_1\right) + b_{H_1} \sin\left(\frac{2\pi j}{N} H_1\right) \tag{22}$$

$$\hat{t}_j = \frac{a_0}{2} + a_{H_1} \cos\left(\frac{2\pi j}{N} H_1\right) + b_{H_1} \sin\left(\frac{2\pi j}{N} H_1\right) + a_{H_2} \cos\left(\frac{2\pi j}{N} H_2\right) + b_{H_2} \sin\left(\frac{2\pi j}{N} H_2\right) \tag{23}$$

We perform our calculations in Microsoft Excel, but we can easily duplicate them using the **FFT** package in R. The results provide the necessary $\hat{y}_i$ to compute the RMSE as outlined by Equation 1.

2. Model Performance

Table 9 highlights the RMSE results using DFT models and includes the estimated parameters, as well as the harmonics identified as the most influential. It also highlights the differences obtained when comparing DFT models and provides confirmation that using a dual harmonic model as compared to using a single harmonic model provides a more accurate data fit across the scenarios we tested. We also concluded that while a single harmonic model was adequate to capture a steady pattern, a dual harmonic model was necessary to model high and low supply due to the presence of near constant inter-arrivals between resupply (Figure 24). Visualization of model fit for the other patterns and random data can be found in Appendix B. Visualization of Model Results.

Table 9. Discrete Fourier Transform Parameters and RMSE Results for Both Single and Double Harmonics.

	Single Harmonic				RMSE	Two Harmonics							RMSE	
Pattern	a0	a1 coef	b1 coef	H1	All Data	a0	a1 coef	b1 coef	a2 coef	b2 coef	H1	H2	All Data	RMSE Differences
Random	105.84	-17.81	-10.77	6	23.4567	105.84	-17.81	-10.772	-18.371	8.337278995	6	3	18.62	4.837
Steady Supply	132.02	-0.369	-38.48	5	37.9565	132.02	-0.369	-38.477	-0.702	30.94124061	5	3	31.012	6.944
High Supply	92.819	1.509	-37.83	2	29.0506	92.819	1.509	-37.83	-26.92	-1.576156819	2	4	21.917	7.134
Low Supply	198.81	27.594	-13.39	2	32.846	198.81	27.594	-13.388	-29.576	7.028485484	2	3	24.835	8.011

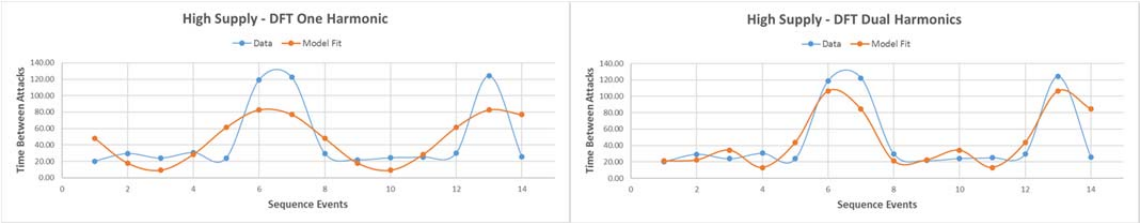


Figure 24. Comparison of Model Fit of High-Supply Pattern Using Single Harmonic vs. Dual Harmonics.

G. COMPARING DFT AND NL OPTIMIZATION

Table 10 compares the RMSE results from DFT modeling to those obtained using the NL optimization methodology. NL optimization outperforms single harmonic models across all patterns and the random data, however, dual harmonic models produce smaller RMSE across all patterns except the steady supply pattern. These results are not unexpected since a single harmonic model uses three terms, NL optimization models use four, and the dual harmonic models use five. The primary benefit to using DFT modeling is the trivial computation run time as compared to the NL optimization methodology.

Table 10. Comparison of RMSE from Single and Dual Harmonic Models to NL Optimization Models

	Pattern	RMSE	NL Optimization RMSE	RMSE Difference
Single Harmonic	Random	23.45672	21.714	-1.743
	Steady Supply	37.9565	26.252	-11.705
	High Supply	29.05057	28.993	-0.058
	Low Supply	32.84605	24.930	-7.916
Dual Harmonic	Random	18.62007		3.094
	Steady Supply	31.0124		-4.761
	High Supply	21.91662		7.076
	Low Supply	24.83526		0.095

The DFT methodology, like NL optimization, produced the lowest RMSEs for the random data. In every variation of DFT testing, the random data produced a lower RMSE suggesting that RMSE would not work well as a metric to signal a possible pattern. This presents a significant issue for this research as we would like to distinguish and highlight IED patterns for further investigation by an intelligence analyst. Rather than labeling NL optimization and DFTs as failed methodologies, Chapter IV will explore the results when the last two IED events of a given data filter are removed as a test set. It will also describe the results of using rolling predictions on the Iraq dataset and highlight the benefits of using DFT models in comparison to naïve models.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. DEVELOPING A TEST SET AND PREDICTING IRAQ IED EVENTS

A. INTRODUCTION

This chapter will begin with an exploration of NL optimization and DFT model performance by predicting the inter-arrival times of the next two IED events after a training sample. We will also perform rolling horizon forecasting of IED activity in Iraq using the all three models (NL Optimization, DFT single and dual harmonic) and compare the results against naïve models.

B. ESTABLISHING A TEST SET AGAINST KNOWN PATTERNS

In order to test the predictive power of our models, we use the first N-2 observations to fit the model and use the last two observations to examine the fit. We refer to this method as the "test-two" method for the remainder of this thesis. Applying this methodology to the test patterns described in Chapter III, only the first 12 inter-arrivals will be used to estimate model parameters. We continue to use root mean-squared error (RMSE) as the metric to compare model performance, but we only use the fit of the last two inter-arrivals in the calculation. Table 11 and 12 contain the results from applying the test-two methodology to the known patterns using the NL optimization model and the DFT model (both single and dual harmonic), respectively. We introduce the four test patterns in Chapter III, and they appear in Figures 17–20. It is important to stress that we only model and test four examples and we therefore cannot draw broad conclusions about the performance of these models across large datasets. However, the results highlight potential issues with pattern recognition using DFT and NL optimization models as well as provide insight into model performance.

Table 11. RMSE of the Last Two Inter-Arrivals Using the NL Optimization Model.

Pattern	Model All Data - RMSE From All	Model 12 Observations - RMSE Last Two Inter-Arrivals					RMSE Difference
		Amplitude (A)	Frequency (f)	PhaseShift (ø)	Offset (K)	RMSE	
Random	21.71	29.834	15.216	16.404	50.771	38.610	16.896
Steady Supply	26.25	55.462	14.663	16.142	72.501	33.283	7.031
High Supply	28.99	-40.851	7.090	23.998	50.271	58.753	29.760
Low Supply	24.93	47.646	19.994	-24.779	95.417	16.456	-8.474

Table 12. RMSE of the Last Two Inter-Arrivals Using Both Single and Double Harmonic Models.

	Pattern	Model All Data - RMSE From All	Model 12 Observations - RMSE Last Two Inter-Arrivals					RMSE Differences			
			a_0	a_1 coef	b_1 coef	H_1	RMSE				
Single Harmonic	Random	23.457	101.450	-28.468	-3.708	5	36.880	13.423			
	Steady Supply	37.956	145.003	-47.693	-28.192	4	32.716	-5.241			
	High Supply	29.051	83.352	-29.643	8.543	1	79.348	50.298			
	Low Supply	32.846	192.560	32.843	27.397	2	15.609	-17.237			
			a_0	a_1 coef	b_1 coef	a_2 coef	b_2 coef	H_1	H_2	RMSE	RMSE Differences
Dual Harmonic	Random	18.620	101.450	-28.468	-3.708	6.269	23.140	5	1	38.013	19.393
	Steady Supply	31.012	145.003	-47.693	-28.192	-14.949	16.121	4	3	45.915	14.903
	High Supply	21.917	83.352	-29.643	8.543	24.062	-13.455	1	2	62.342	40.425
	Low Supply	24.835	192.560	32.843	27.397	2.442	-28.085	2	4	9.111	-15.724

The tables include the estimated parameters for the models using the test-two methodology as well as the RMSE results from both the original model and the new approach. The column labeled “Model all Data” contains the RMSE results from Chapter III where each model was fit, and the RMSE was calculated, across all 14 observations. The RMSE column contains the calculated results from using the test-two methodology and the RMSE difference (last column) highlights the changes in RMSE between the methodology used in Chapter III and the test-two methodology.

One of the primary issues with both the NL optimization and DFT models, as described in Chapter III, is their inability to distinguish between random data and known patterns. Using the test-two methodology, the DFT single harmonic and the NL optimization models may provide some delineation between pattern and random data. All three methodologies produce higher RMSE modeling the random data using the test-two methodology (Table 1) than in the base case fitting all 14 observations in sample. This suggests that all three models are poor

at predicting random IED activity. Additionally, the DFT single harmonic model produces lower RMSE for the steady supply patterns and low supply patterns. When modeling the steady supply pattern, both the DFT single harmonic and NL optimization models produce better results than when they model random data. The poor performance of the DFT dual harmonic model may be the product of over-fitting. We next explore why the fit for the RMSE for the high supply model increases so much when we use the test-two approach.

The high supply dataset experiences a RMSE increase double to quadruple (depending on the model) that of the random sequence. Comparing the data curves of the low supply pattern versus the high supply pattern as highlighted by Figures 25 and 26, respectively, provides a clear distinction in terms of the number of cycles present for modeling. The low supply pattern includes two full cycles while the high supply pattern only provides one and one-half for the first 12 modeled observations.

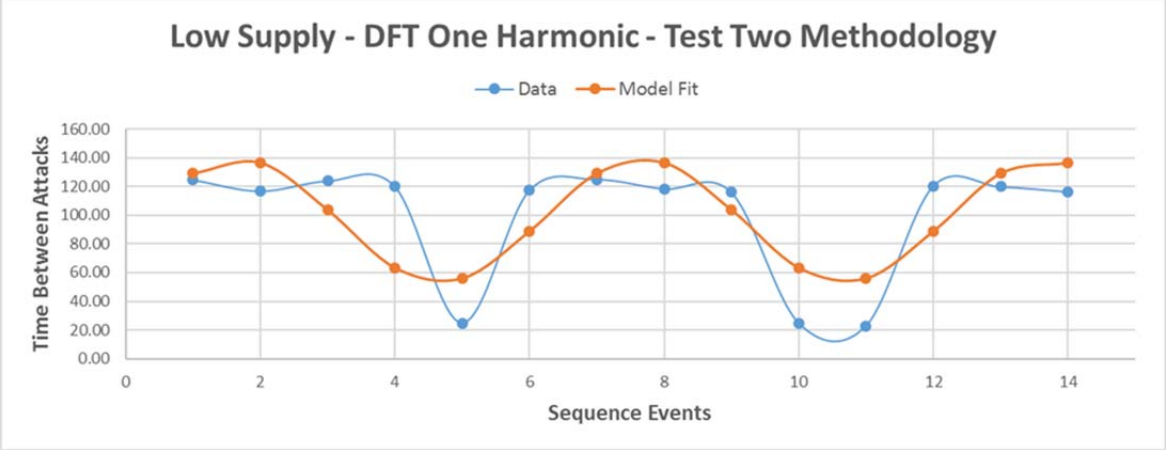


Figure 25. DFT Single Harmonic Model Fit of Low Supply Pattern using Test Two Methodology.

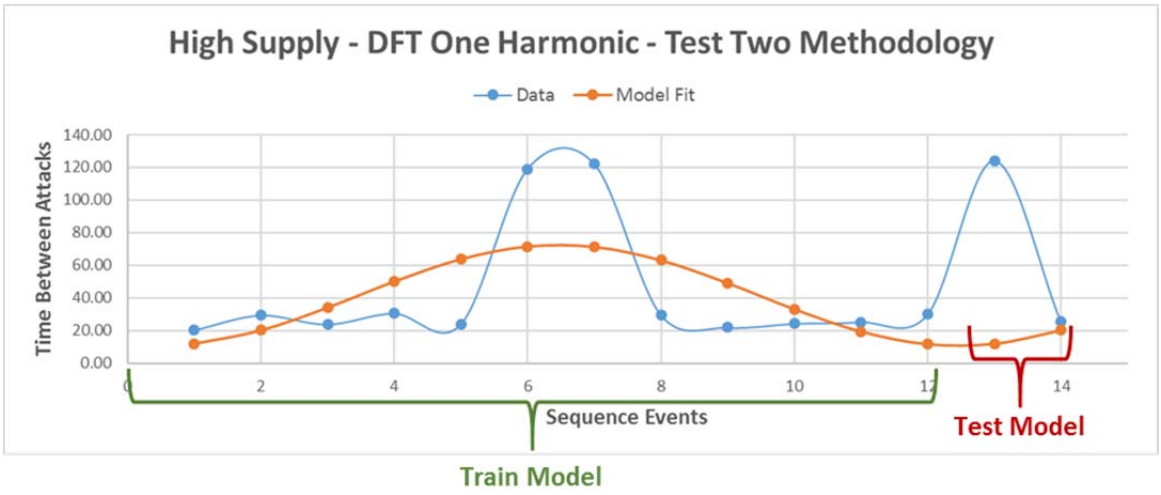


Figure 26. DFT Single Harmonic Model Fit of High Supply Pattern using Test Two Methodology.

The DFT approach will produce a pattern that repeats periodically due to its sinusoidal form. Therefore, when estimating parameters for the DFT, it is better if the underlying data correspond to roughly an integer number of periods or cycles. For example, when we remove the last two observations from the low supply pattern, the remaining 12 observations form essentially two complete cycles and thus the DFT model produces reasonable results (Figure 25). The two complete transitions from long inter-arrivals to short inter-arrivals and back allows the model to identify the rough number of observations between large transitions. However, when we remove the last two observations from the high supply model in Figure 26, the remaining observations only form approximately 1.5 cycles, which prevents accurate modeling of long inter-arrivals. Figure 27 represents the same high supply pattern with two additional observations consistent with the pattern. Now that the training set constitutes two full cycles, the RMSE drops from 79.3 to 16.8.

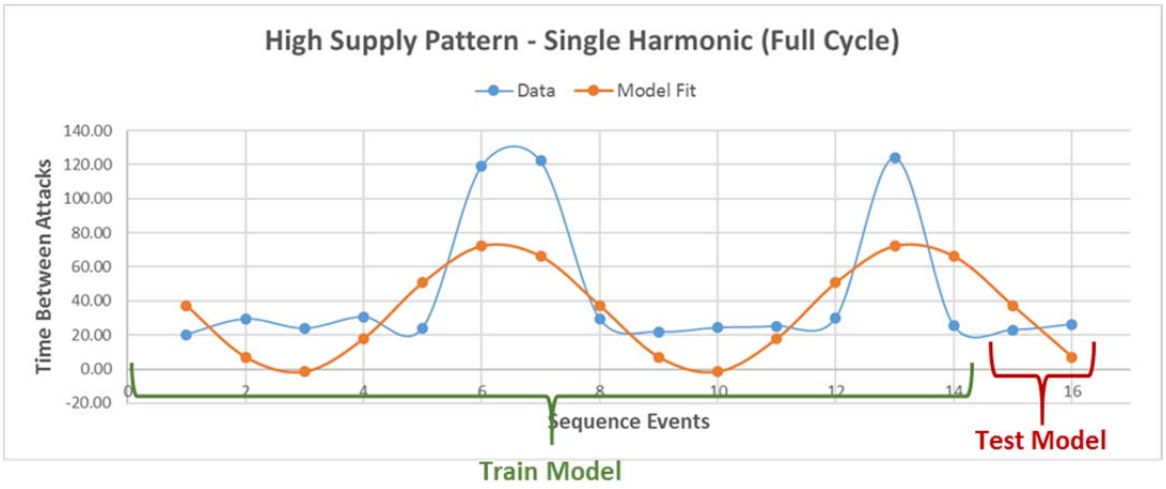


Figure 27. Full Cycle DFT Single Harmonic Model Fit of High Supply Pattern using Test Two Methodology.

Modeling of the steady pattern using the test-two methodology produced varied results. Results degraded using both the NL optimization and DFT dual harmonic models (most likely the result of over-fitting) but improved using the DFT single harmonic model. The DFT single harmonic model did improve slightly using the test-two methodology (Figure 4) and it out performed the fit of the random data but it failed to produce an RMSE similar to the low supply pattern and high supply pattern after observations were added (RMSE<20). While Figure 28 illustrates the DFT single harmonic model captures the oscillations reasonably well at a high level, there are significant errors in the model fit for certain observations. Unfortunately, accurate modeling of the steady supply pattern proves difficult for reasons similar to periodicity problems we saw with the high supply model, as well as the existence of variability in the number of observations before transitions between short and long inter-arrivals. These challenges limit the situations where DFT and NL optimization models perform well fitting steady supply patterns. Figures 29 and 30 provide a visual representation of situations where DFT models accurately portray steady supply patterns. Unsurprisingly, these examples highlight the need for consistent sinusoidal data with very low variability. The pattern in Figure 29 is consistent in

terms of the number of inter-arrivals between transitions (two short inter-arrivals and one long inter-arrivals) while Figure 30 confirms that accurate modeling of steady supply includes patterns with multiple arrivals between transitions.

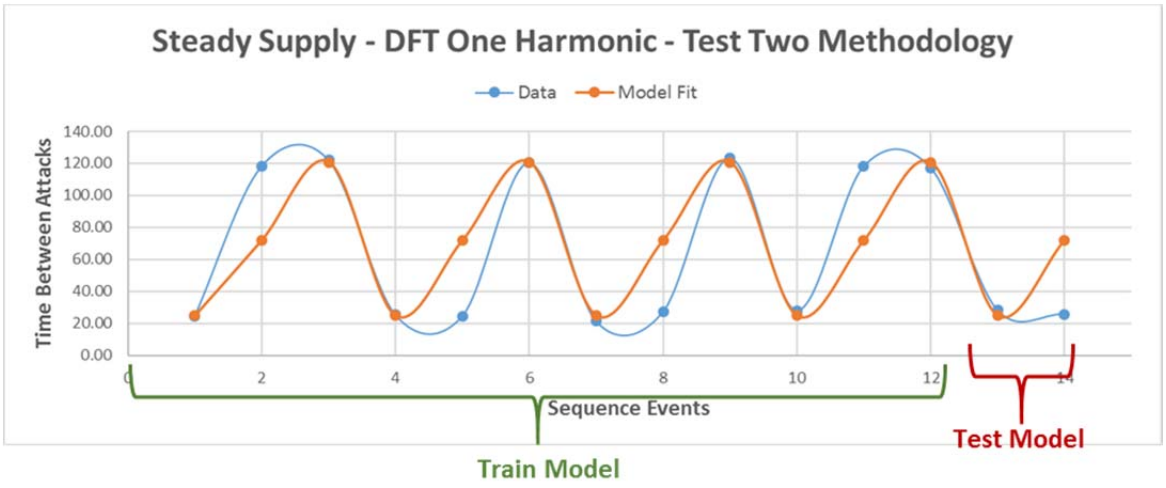


Figure 28. DFT Single Harmonic Fit of Steady Supply Pattern Using Test-Two Methodology.

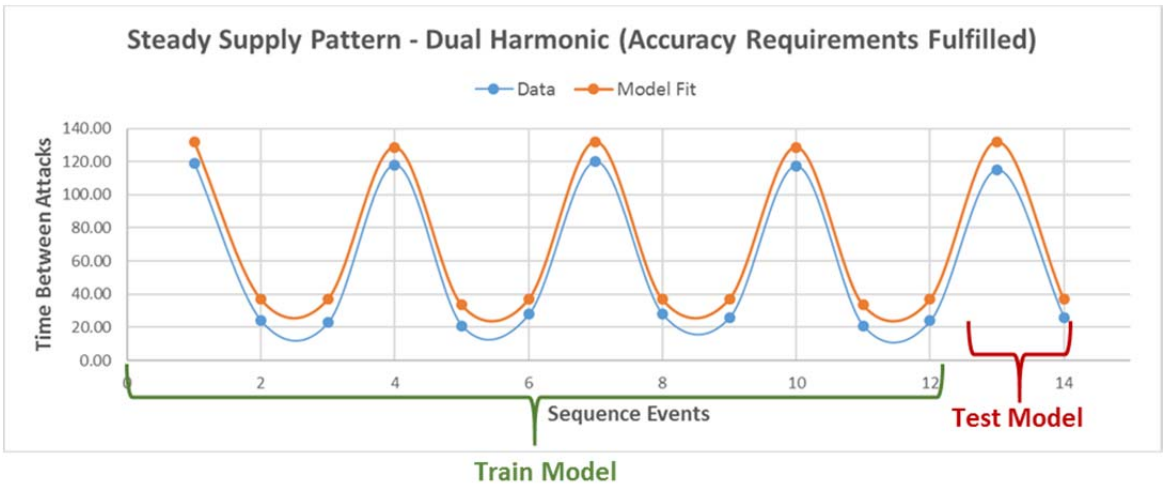


Figure 29. Accurate Model Fit of Consistent Steady Supply Pattern (Near Perfect Sinusoid).

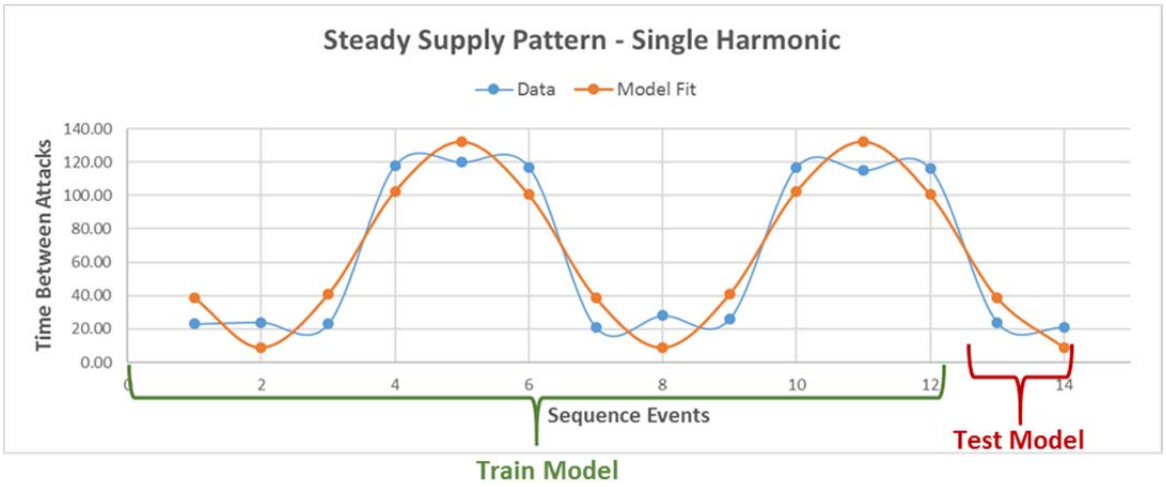


Figure 30. Accurate Model Fit of Steady Supply Pattern with Additional Inter-Arrivals before Transition

Table 13 summarizes the results from Tables 11 and 12. It also provides a comparison of the RMSE results from our models against a new model that uses the mean inter-arrival of the first 12 observations to predict the last two observations. The mean inter-arrival model should perform better when modeling random inter-arrivals and limitless supply (visual flat line) as compared to the sinusoidal patterns we test. The results are as expected with the mean inter-arrival model performing best against the random data set with the other models lagging far behind. The steady supply, high supply (after completing the cycle as previously discussed), and low supply are all modeled more accurately using the DFT single harmonic and NL optimization. The only unexpected result was the outperformance of the dual harmonic model by the mean inter-arrival against the steady supply pattern, which can be explained by over-fitting.

Table 13. Model RMSE Comparison against Mean Inter-arrivals

Pattern	RMSE using Mean Inter-Arrival	RMSE - Single Harmonic	RMSE - Dual Harmonic	RMSE - NL Optimization
Random	16.66	36.88	38.01	38.61
Steady Supply	45.46	32.72	45.92	33.28
High Supply	59.41	79.35	62.34	58.75
High Supply - After Cycle Completion	21.89	16.87	5..84	15.9
Low Supply	21.93	15.61	9.11	16.46

The results of the test-two methodology suggest there may be some potential to delineate between pattern and randomness when using NL optimization and DFT single harmonic models. If the mean inter-arrival fits much better than the NL or DFT models, the observations probably do not form a legitimate pattern. The analysis also highlights some significant obstacles to accurate model fits. Chief among them are the need for complete cycles for high and low supply patterns as well as the need for consistent data when modeling steady supply patterns. Additional model testing is necessary and we will evaluate model performance after using the test-two methodology to model portions of the Iraq dataset.

C. ROLLING HORIZON FORECASTING OF THE IRAQ DATASET

To further test our models, we chose to perform rolling horizon forecasting to compare model performance against the mean inter-arrival rate across large subsets of Iraq data (Hyndman et al. 2006). This testing methodology allows us evaluate model performance with an “auto-pilot” approach. The first step is to filter the data in both space and time, which is followed by rolling predictions as we step through the data.

We temporally filtered the data into 6-month blocks to ensure we capture enough data at a local level. The 6-month window also allows us to capture IED activity across transitional periods such as the “Sons of Iraq” movement, which could potentially affect model performance. We choose three testing time windows: September ’05 to February ’06, August ’06 to January ’07, and July ’07 to December ‘07 (Figure 31). We next filter spatially based on natural spatial

breaks between IED clusters to narrow our focus (Figure 32). The filtering process produces eight spatio-temporal data subsets covering three 6-month windows over five different areas of Iraq (some subsets were of the same area but with a different temporal filter). The spatio-temporal subsets range from 114 IEDs to 244 IEDs for a combined total of 1420 IEDs across all eight subsets.

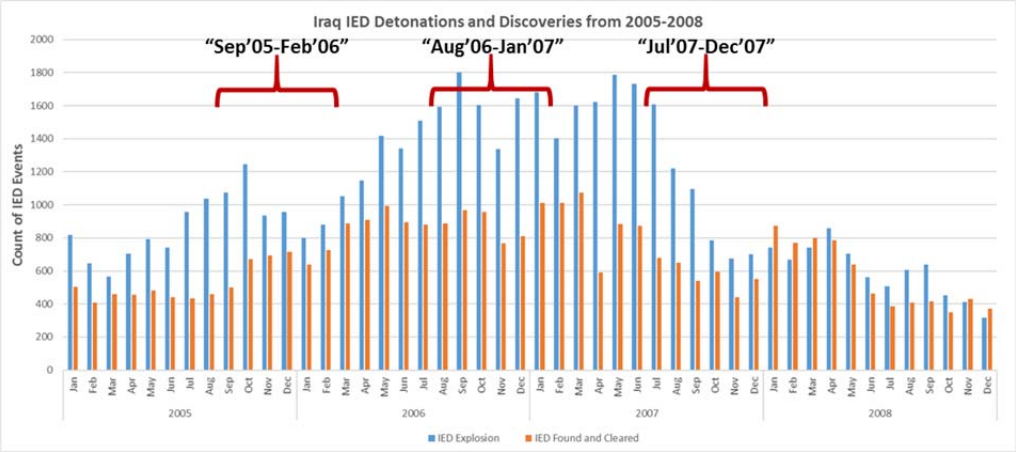


Figure 31. Temporal Subsets for Rolling Horizon Forecasting.

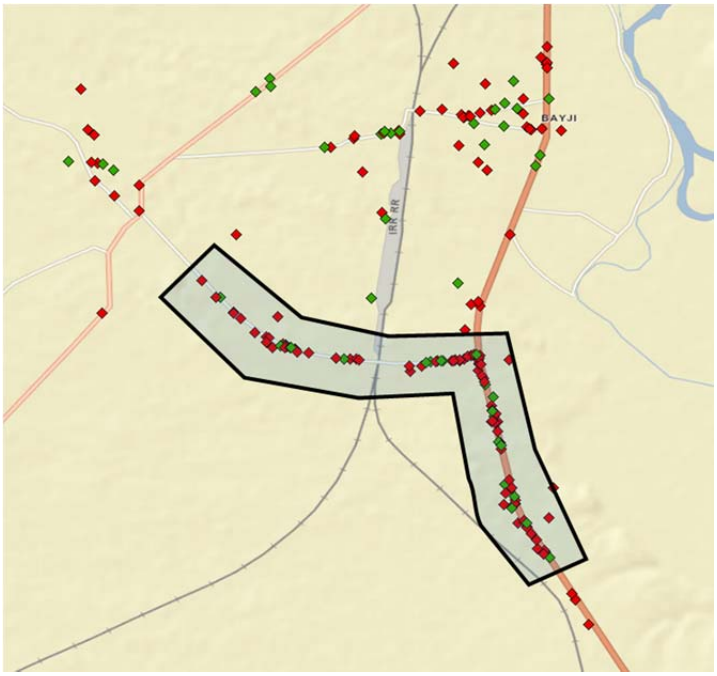


Figure 32. Spatial Filtering for Rolling Horizon Forecasting (9.1 km).

Rolling horizon forecasting is a time step methodology wherein we shift the data window of interest forward at each step and consider a slightly different sequence from the previous step. In this case, we use rolling horizon forecasting in conjunction with the test-two methodology previously described. We start by examining the inter-arrival times *1* through *m*. We fit a model to these m observations and then predict the inter-arrival times *m+1* and *m+2*. Based on the predictions of these two observations, we compute the RMSE. Here, *m* represents an appropriate number of observations to model (between 6 and 23 as described in Chapter II). After the prediction of inter-arrival times *m+1* and *m+2*, the rolling horizon forecast steps forward a single observation, fits the model to observations *2* through *m+1*, and predicts observations *m+2* and *m+3*. (Figure 33). We continue this process throughout the dataset producing N-*m*-1 RMSE values for each model (N is the number of inter-arrivals in a subset).

The final step is to choose the number of observations for modeling and testing. We perform this analysis for varying observation lengths and produce similar results across the eight 6-month subsets. For our purposes, we choose to present the results of a modeling window of 14 total observations: 12 for model development and the last two observations to calculate RMSE values. These numbers are consistent with our previous models. We may obtain better results by varying the number of modeled observations depending on the geographic area or the number of inter-arrivals in a subset.

Step	Observations															
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
		Observation for Modeling								Observation for Predicting							

Figure 33. Visual Representation of Rolling Horizon Forecasting.

Table 14 summarizes the results of the rolling horizon forecasting across all eight spatio-temporal subsets producing 1328 RMSE values per model (DFT single and dual, NL optimization, mean inter-arrival). The mean inter-arrival model produces the lowest RMSE 42 percent of the time, DFT single and dual harmonic 16 and 17 percent of the time, respectively, and the NL optimization model is the best performer 24 percent of the time. The better performance of the mean inter-arrival model is not surprising, as a majority of the IED sequences modeled do not constitute a legitimate sinusoidal pattern. As previously discussed, the mean inter-arrival performs better in situations with un-patterned data. The last row in Table 14 measures how often that particular model performed better than the mean inter-arrival. The similarity between these numbers is also expected since combinations of IEDs better modeled by a single sinusoidal model are likely to also be better modeled by multiple sinusoidal models when compared to the mean inter-arrival.

Table 14. Summary of Model Performance using Rolling Horizon Forecasting.

N=1328	DFT	DFT2	NL Optim	Mean
Best Model	218	231	316	561
Best Model Percentage	16.4%	17.4%	23.8%	42.2%
Percentage Model is Better than using Mean Inter-Arrival	32.3%	32.8%	33.3%	

D. TESTING CANDIDATE PATTERNS FROM THE IRAQ DATA

The results of our rolling horizon forecast suggest that if we apply these methodologies blindly across an entire dataset, the mean inter-arrival model out-performs any of our sinusoidal models. Without further analysis, we could conclude that the mean inter-arrival is the most predictive of the four and should be used in all situations. We next focus on only a small number of sequences that appear to be patterns. We examine the model performance against these potential patterns from the Iraq data and compare the results to those obtained

using the rolling horizon analysis for the entire dataset discussed in the previous section.

To accomplish this, we inspected the inter-arrival curves for all 1328 IED sequences tested during the rolling horizon forecast to identify likely candidate patterns. We select only the curves that, in my personal experience, would warrant a briefing to my commander and a request for assets. We perform this selection process without consulting any models or additional metrics; we only performed a visual inspection. We identify 19 patterns, which highlights how rare IED pattern recognition is in a deployed environment. Figures 34 and 35 are examples of the curves we identify as patterns, which have distinct sinusoidal shapes consistent with the steady and high supply patterns, respectively.

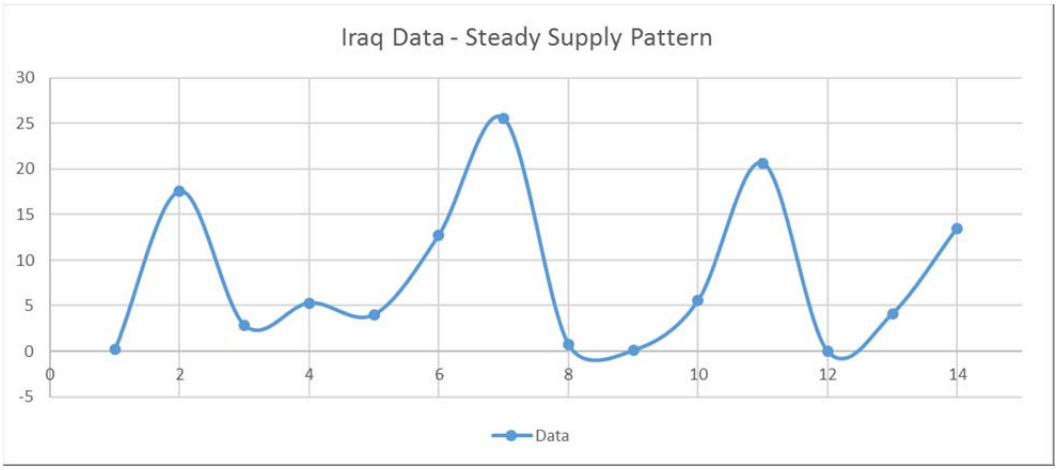


Figure 34. Example of Steady Supply Pattern from the Iraq Data.

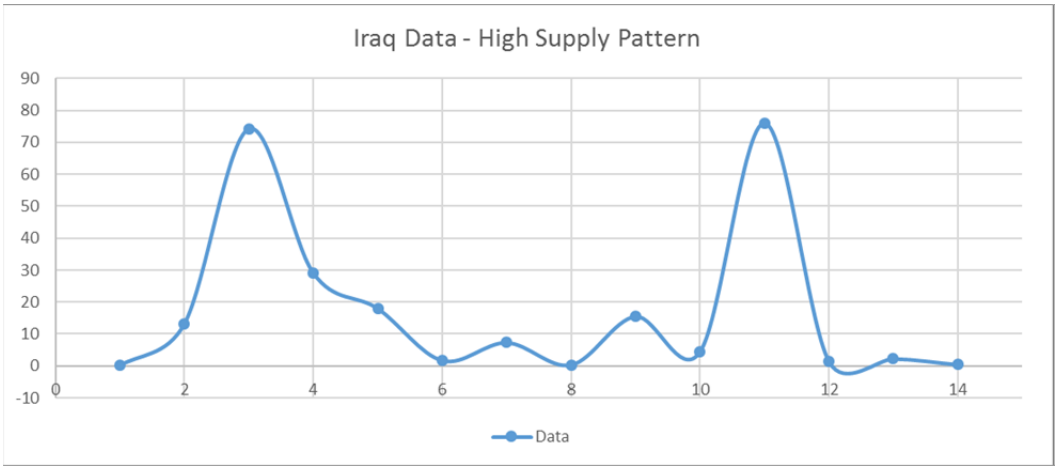


Figure 35. Example of High Supply Pattern from the Iraq Data.

Any pattern identification process (such as this visual inspection procedure) has false positives and false negatives. False positives are IED sequences that we identify as patterns that are not actually patterns. A related issue is legitimate patterns that are disrupted before the next event and do not continue into the future. From an operational point of view, false positives and disrupted patterns are equivalent, as we are unlikely to deploy assets in an effective manner to interdict the next IED based on these observations. Anecdotal evidence suggests 40–50 percent of patterns identified by visual inspection are either false positives or disrupted patterns. False negatives are patterns that are not identified, often due to the existence of IED noise, and/or incorrect filtering of space, time, or type of IED device. Table 15 has the same structure as Table 14 and provides summary results for fitting these identified potential patterns with our models. The results are significantly different than the results from our rolling horizon analysis in Table 14.

Table 15. Summary of Model Performance against Confirmed Iraq Patterns

N=19	DFT	DFT2	NL Optim	Mean
Best Model	6	2	11	0
Best Model Percentage	31.6%	10.5%	57.9%	0.0%
Percentage Model is Better than using Mean Inter-Arrival	84.2%	57.9%	89.5%	

Of the 19 identified patterns, the mean inter-arrival model is never the best model to use. The DFT Double Harmonic model performs the best 58 percent of the time. The performance of the DFT single harmonic model and NL optimization models in comparison to the mean is significant. They out-performed (produced lower RMSE) 84 and 90 percent of the time, respectively. Using this set of data as an example, the current pattern detection approach would require the analyst to inspect all 1328 sequences visually (our first step in this process). These results suggest that analysts attempting to find patterns can significantly improve search efficiency if they start with IED combinations where the DFT single harmonic or the NL optimization models perform better than the mean inter-arrival. Since DFT single harmonic and NL optimization out-perform mean inter-arrival roughly 30 percent of the time individually (see Table 14), analysts could reduce workload by 70 percent. Analysts narrowing their focus to sequences where NL optimization or DFT single harmonic perform better than the mean inter-arrival would still need to go through nearly 400 sequences (1328*0.3). However, these 400 would contain 90 percent (using NL optimization) and 84 percent (using DFT) of the identified patterns in this example. Figures 36 and 37 provide a visual representation of two of the 19 identified patterns as well as the fit of all four models. Further examples appear in Appendix C. Model Fits of Candidate Iraq Patterns.

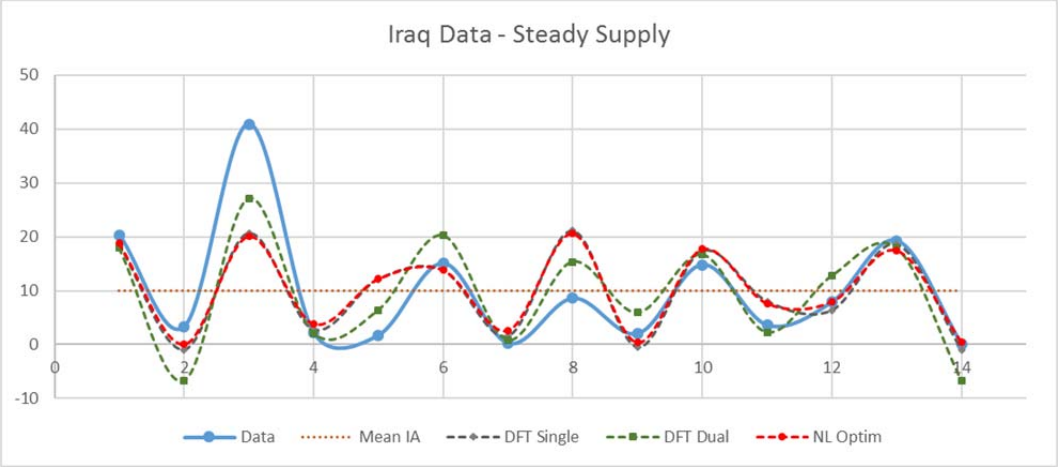


Figure 36. Model fits of Confirmed Steady Supply Pattern from the Iraq Dataset.

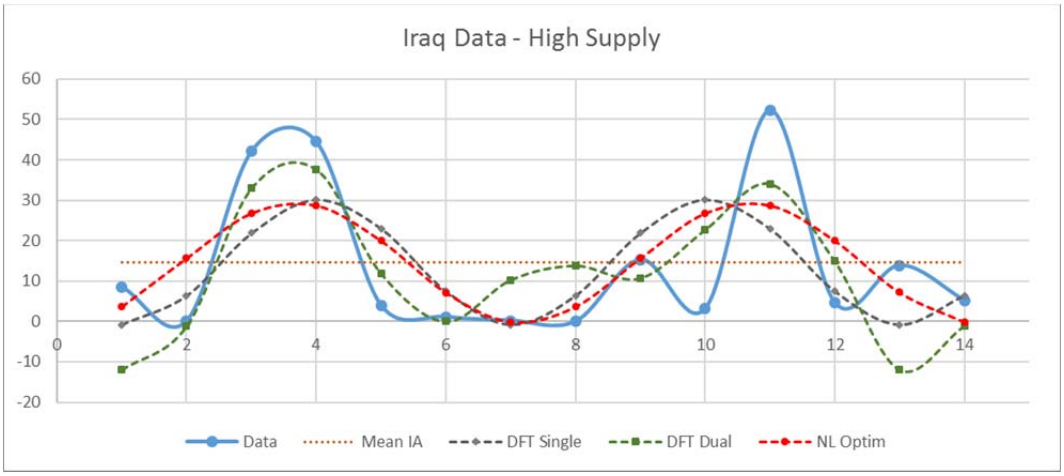


Figure 37. Model fits of Confirmed High Supply Pattern from the Iraq Dataset.

Our final analysis evaluates model performance when predicting the next event of an identified pattern. In most situations, analysts spend a majority of time attempting to identify patterns. When an analyst identifies a pattern, it is common practice to determine the most likely time for the next event by visually inspecting the inter-arrival curve in combination with historic day of week and time of day analysis for the local area. This type of analysis might also consider

type of device or initiation. Adding this additional analysis helps reduce the false positive rate by eliminating illogical patterns and predictions. In practice, analysts would use more than just the inter-arrival time information to predict the future IED events. However, we are curious to see how our models performs with only historic inter-arrival times as input.

The 19 patterns we identify in the Iraq dataset are the baseline in this analysis. Since they are likely pattern candidates, we develop the model based on all 14 observations and predict the 15th. Table 16 provides the results and the mean inter-arrival model performs the best when we consider all 19 patterns. These results are not surprising since our identified patterns include sequences where the 15th observation deviates significantly from previous pattern behavior. Figure 38 provides an example visual representation of a candidate pattern with a deviation in the 15th observation. The red dot represents a consistent inter-arrival value if the pattern persists past the 14th observation. If we examine the results of only the patterns that persist through the 15th observation, the NL optimization model produces a lower RMSE than the mean inter-arrival (Figure 39). Relying on personal experience, 11 of 19 of the candidate patterns meet this criterion based on visual inspection of the sinusoidal pattern established with the first 14 observations.

Table 16. RMSE from Predicting the 15th Event of Iraq Pattern Candidates.

N=19	DFT	DFT2	NL Optim	Mean
RMSE Results - All 19	21.62	24.16	18.89	13.17
RMSE Results - Only Persistent Patterns	13.61	12.82	5.86	8.12

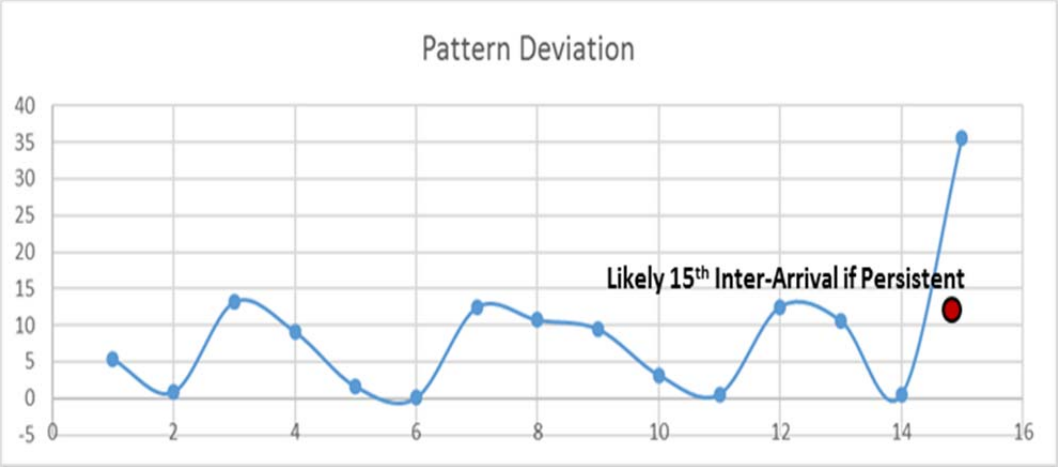


Figure 38. Pattern Deviation of the 15th Observation from Iraq Pattern Candidates.

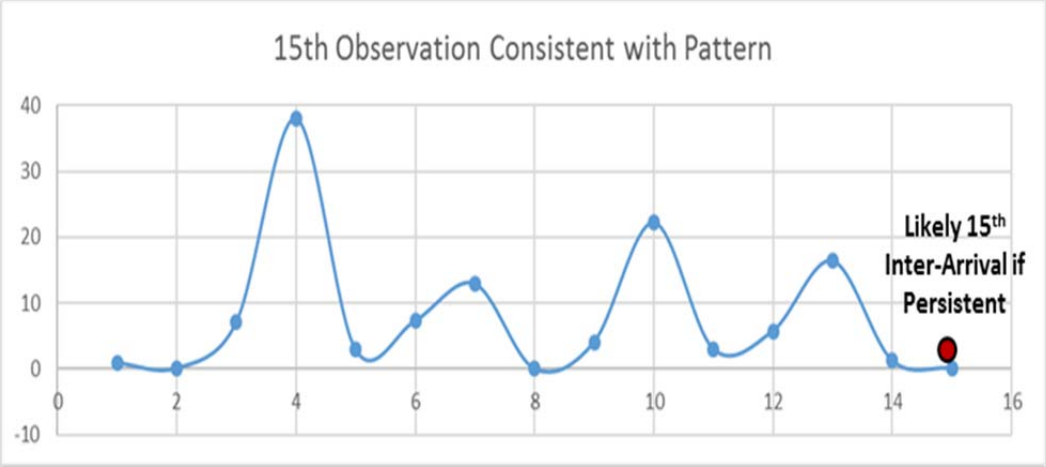


Figure 39. Pattern Consistency with the 15th Observation from Iraq Pattern Candidates.

Even though we cherry-pick the results in the 2nd row of Table 16, it highlights why our methodology may have some operational relevance. All 19 of our identified patterns are strong enough candidates to apply either route clearance or surveillance assets with the goal of finding the device or disrupting the emplacement. In a deployed environment, the analyst only has access to the first 14 events of any of these patterns and his "best guess" for the next event is going to be a continuation of the current pattern. If the pattern does not continue,

either because it is a false positive or it was disrupted, then it is unlikely any method (including the mean) will accurately predict the timing of the next event. This appears to be the case in 8 of the 19 potential patterns, which is consistent with my experience that 40–50 percent of visually identified patterns do not continue. However, if the pattern persists, our results suggest that using a NL optimization model provides reasonable prediction of the 15th event in the absence of additional information. From an operational point of view, this is important as our methods perform better when the pattern persists and will perform similarly (i.e., poorly) when the pattern does not continue when compared to the mean inter-arrival.

V. CONCLUSION

IEDs provided a significant threat to U.S. and coalition forces in Iraq and Afghanistan. Their success against superior forces ensures their continued use as the insurgent weapon of choice. Intelligence analysts attempting to recognize and identify IED patterns at a local level have a difficult task. There is no “silver bullet” to IED pattern recognition and prediction, but it is crucial that IED analysts take advantage of every insurgent mistake in the form of predictable patterns. A single IED cell can emplace hundreds of IEDs before presenting a predictable pattern. In my 24 months of deployment time as an intelligence analyst for a route clearance battalion, my team successfully identified and disrupted only four recognized patterns. Finding a better, more efficient, way to recognize IED patterns is necessary.

While this thesis unfortunately did not produce results that would identify patterns with low false positives and false negatives, the methods and models discussed in this thesis have potential to reduce IED analyst workload. Analysts do not have time to continuously search for patterns with competing requirements. Our results suggest that an analyst who focuses on IED sequences where the NL optimization or the DFT single harmonic model out-performs the mean inter-arrival model may be able to reduce the amount of sequences analyzed while still considering most of the visually recognizable patterns. When we applied our methodology, the workload was reduced by 70 percent but the remaining sequences contained 90 percent of the candidate patterns. This requires the analyst to evaluate nearly 400 IED sequences; a formidable task, but much more easily accomplished than the evaluation of the original 1328 IED sequences. Additionally, we establish that NL optimization models predict the inter-arrival of the next observation fairly well in situations where an identified pattern persists. This is a useful operational result.

We also identify and discuss the shortfalls of our models. We initially focused on the Hawkes point process, but did not pursue it because of its

inability to accurately model small datasets. We examined sinusoidal models such as DFT and NL optimization. Sinusoidal models require data to be in near full cycles, which considerably increases the importance of selecting the right number of observations. As an example, our modeling of the high supply pattern produced poor results until we added two more observations, which completed a cycle. We found that, in general, the DFT dual harmonic model is prone to overfitting and produced the worst results of the models tested. Lastly, our NL optimization model arguably performed the best, in both pattern delineation and IED prediction, but requires considerable run times and would not be suitable for datasets exceeding several thousand sequences. Most tactical level modeling performed on a daily/weekly basis would only require an analyst to consider usually well less than 100 observations. However, more in depth analysis would remove certain observations and only focus on a subset of the original dataset. For example, if the original data set has only 24 inter-arrival times, the analyst might want to consider subsets of size 15 and evaluate whether any of these subsets form patterns. The analysis quickly becomes computationally burdensome when we consider subsets.

We conclude that even though the DFT single harmonic model performs slightly worse than the NL optimization, its run times are trivial making it the model of choice when it is necessary to examine large numbers of sequences. As discussed above, a potential topic for future work is the exploration of how to generate and evaluate sequences that are subsets of the dataset to maximize pattern identification while minimizing false positives and false negatives. This may involve additional variables such as initiation or explosive type. An analyst may want to account for the possibility of IED noise by ignoring certain observations and focusing on non-continuous subsets of the most recent IED activity. We discuss IED noise in Chapter II, but do not incorporate it into our analysis in Chapters III and IV. As discussed in Chapter III, evaluating subsets is a computationally expensive process since it is necessary to check every possible sequence of length 6,7,8, etc. for a given dataset. Unfortunately, this will

produce millions (if not many more) of sequences to consider and will likely result in a large number of false positives. Future research can examine more sophisticated ways to subset the data. One possibility is to start the analysis with a small subset and grow that subset by adding the observations most likely to accentuate an existing pattern or highlight a new pattern not previously seen.

There are other opportunities for future work. Our results suggest that sinusoidal models perform best with complete cycles, and determining the number of events to model is crucial to the identification of possible patterns. Future research could explore the possibility of a dynamic approach where an analyst begins with a fixed number of observations but the model is free to include or exclude a set number of inter-arrivals to improve the pattern. It may be worthwhile to determine if these models perform better or worse in urban versus rural environments. There tend to be many more insurgent groups operating in close proximity in urban environments. Sometimes these groups are coordinated, but often they are not, which makes uncovering IED patterns more difficult. Lastly, it is vital that the NL optimization and DFT single harmonic model are implemented into a VBA based tool for deployed analysts. Analysts have very few tools at their disposal to visualize IED patterns, and none that can help focus their efforts by eliminating possible sequences. Analysts often create their own searching mechanism but a tool with these models built in has the potential to benefit units searching for IED patterns.

The objective of this thesis is to explore methodologies to help deployed analyst identify IED patterns. Although we do not develop a single method that provides the analyst with all the solutions, we formulate a particular set of models with the potential to decrease the amount of time required to identify patterns.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. SCRIPTS

A. R SCRIPT FOR THE HAWKES POINT PROCESS

```
###Library###
library(hawkes)

###Function to determine log-likelihood###
neg.loglik <- function(params, data, opt=TRUE) {

  #baseline rate that events occur
  mu <- params[1]
  #immediate jump value
  alpha <- params[2]
  #decay rate
  beta <- params[3]
  #data is time of the events
  t <- sort(data)

  #r is intermediate value need to define recursively
  r <- rep(0,length(t))
  for(i in 2:length(t)) {
    r[i] <- exp(-beta*(t[i]-t[i-1]))*(1+r[i-1])
  }

  loglik <- -tail(t,1)*mu
  loglik <- loglik+alpha/beta*sum(exp(-beta*(tail(t,1)-t))-1)
  loglik <- loglik+sum(log(mu+alpha*r))
  if(!opt) {
    return(list(negloglik=-loglik, mu=mu, alpha=alpha, beta=beta, t=t,
                r=r))
  }
  else {
    return(-loglik)
  }
}

######Read in Data######
data= read.csv("Low_Supply.csv",header=FALSE)
tVec=as.numeric(data$V1)

RMSE=1000
best.value=1000
for (j in 1:1000) {
       # Determine values for (mu, alpha, beta) using loglik function
       opt       <-       optim(par=c(runif(1,0,10),runif(1,0,10),runif(1,0,10)),
fn=neg.loglik,  data=tVec,  lower  =  c(0.0000001,0.00000001,0.0000001),   method  =
"L-BFGS-B")
       opt$par
       if (opt$par[3]>100) next
       if (opt$par[2]>opt$par[3]) next
       #Simulate hawkes 1000 times using calculated parameters and combine into
dataframe
       sim.mat=as.matrix(tVec)
       if (opt$value<best.value) {
              best.value=opt$value
              best.para=opt$par
              }
  }
```

```
for (i in 1:1000) {
       history<-simulateHawkes(best.para[1],best.para[2],best.para[3],3000)
       sim=as.numeric(history[[1]])
       sim=sim[1:length(tVec)]
       sim.mat=cbind(sim.mat,sim)
       }

#Conversions to inter-arrivals for RMSE calculations
tVec.ia=c(tVec[1], as.numeric(unlist(lapply(data,diff))))
sim.mat=data.frame(sim.mat)
sim.ia=as.matrix(tVec.ia)
for (i in 2:1001){
       temp.df=data.frame(sim.mat[,i])
       temp.vec=c(temp.df[1,1], as.numeric(unlist(lapply(temp.df,diff))))
       sim.ia=cbind(sim.ia,temp.vec)
       }

#Calculation of RMSE for every simulation
rmse.vec=numeric(0)
for (i in 2:1001){
       col.rmse=rmse(sim.ia[,1],sim.ia[,i])
       rmse.vec[i-1]=col.rmse
       }

#calculate RMSE Average
rmse.mean=mean(rmse.vec)
```

B. R SCRIPT FOR RUNNING HORIZON FORECAST (MEAN, DFT, DFT2)

```
#Library
library("Metrics")

#Read in the data
nai.df=read.csv('Virginia_sep05-feb06.csv',header=T,stringsAsFactors=F)

#Clean the data
dt.vec=strptime(nai.df$Date_Time, format='%m/%d/%y %H:%M')
ia.vec=difftime(tail(dt.vec,-1), head(dt.vec,-1),units="hours")
ia.vec=as.numeric(ia.vec)
N=length(ia.vec)


#########Mean Inter-arrival#########
#Initialize empty list and vector
m.rmse.list=list()
m.rmse.mean=numeric()
#Loop through sequence lengths
for (event.count in 7:22) {
       #Initialize and empty vector after each length
       temp.vec=numeric()
       #Loop through events for running prediction and capture the RMSE from two
predicted events
       for (j in 1:(N-event.count-1)) {
              temp.mean=mean(ia.vec[j:(event.count+j-1)])
              temp.pred=rep(temp.mean,2)

       temp.rmse=rmse(ia.vec[(j+event.count):(j+event.count+1)],temp.pred)
              temp.vec[j]=temp.rmse
       }
```

```
	#populate a list with the RMSE
	m.rmse.list[event.count]=list(temp.vec)
	#calculate the mean RMSE for a single sequence length (7:24 at the end)
	m.rmse.mean[event.count-6]=mean(m.rmse.list[[event.count]])
}
names(m.rmse.list)=c("one","two","three","four","five","six","seven","eight","n
ine","ten","eleven","twelve","thirteen","fourteen","fifteen","sixteen","sevente
en","eighteen","nineteen","twenty","twentyone","twentytwo")
m.rmse.list[c("one","two","three","four","five","six")]=NULL


#########Last Inter-arrival#########
LIA.rmse.list=list()
LIA.rmse.mean=numeric()
for (event.count in 7:22) {
	temp.vec=numeric()
	for (j in 1:(N-event.count-1)) {
		temp.LIA=ia.vec[(j+event.count-1)]
		temp.pred=rep(temp.LIA,2)

	temp.rmse=rmse(ia.vec[(j+event.count):(j+event.count+1)],temp.pred)
		temp.vec[j]=temp.rmse
	}
	LIA.rmse.list[event.count]=list(temp.vec)
	LIA.rmse.mean[event.count-6]=mean(LIA.rmse.list[[event.count]])
}
names(LIA.rmse.list)=c("one","two","three","four","five","six","seven","eight",
"nine","ten","eleven","twelve","thirteen","fourteen","fifteen","sixteen","seven
teen","eighteen","nineteen","twenty","twentyone","twentytwo")
LIA.rmse.list[c("one","two","three","four","five","six")]=NULL


#########Discrete Fourier Transforms (single and double harmonic
model)#########
DFT.rmse.list=list()
DFT.rmse.mean=numeric()
DFT2.rmse.list=list()
DFT2.rmse.mean=numeric()
for (event.count in 7:22) {
	temp.vec=numeric()
	temp.vec2=numeric()
	for (j in 1:(N-event.count-1)) {
		tVec=ia.vec[j:(event.count+j-1)]
		M=length(tVec)
		#get fourier transform from built in R functions
		yk = fft(tVec)/length(tVec)
		Avec = rep(0,1+floor((M-1)/2))
		Avec = 2*Re(yk[1:length(Avec)])
		#get the coefficient for B
		Bvec = -2*Im(yk[1:length(Avec)])
		sqrCoefficients = Avec[2:length(Avec)]^2+Bvec[2:length(Avec)]^2
		I = which.max(sqrCoefficients )
		sqrCoefficients[I]=0
		I2 = which.max(sqrCoefficients)
		ffBestHarmonic  =  Avec[1]/2+  Avec[I+1]*cos(2*pi*I*(0:(M+1))/M)  +
Bvec[I+1]*sin(2*pi*I*(0:(M+1))/M)
		ffBest2    =    Avec[1]/2+    Avec[I+1]*cos(2*pi*I*(0:(M+1))/M)    +
Bvec[I+1]*sin(2*pi*I*(0:(M+1))/M) +
                                Avec[I2+1]*cos(2*pi*I2*(0:(M+1))/M)          +
Bvec[I2+1]*sin(2*pi*I2*(0:(M+1))/M)
		temp.pred=tail(ffBestHarmonic,2)
		temp.pred2=tail(ffBest2,2)
```

```
        temp.rmse=rmse(ia.vec[(j+event.count):(j+event.count+1)],temp.pred)

        temp.rmse2=rmse(ia.vec[(j+event.count):(j+event.count+1)],temp.pred2)
               temp.vec[j]=temp.rmse
               temp.vec2[j]=temp.rmse2
        }
        DFT.rmse.list[event.count]=list(temp.vec)
        DFT.rmse.mean[event.count-6]=mean(DFT.rmse.list[[event.count]])
        DFT2.rmse.list[event.count]=list(temp.vec2)
        DFT2.rmse.mean[event.count-6]=mean(DFT2.rmse.list[[event.count]])

}
names(DFT.rmse.list)=c("one","two","three","four","five","six","seven","eight",
"nine","ten","eleven","twelve","thirteen","fourteen","fifteen","sixteen","seven
teen","eighteen","nineteen","twenty",twentyone","twentytwo")
DFT.rmse.list[c("one","two","three","four","five","six")]=NULL
names(DFT2.rmse.list)=c("one","two","three","four","five","six","seven","eight"
,"nine","ten","eleven","twelve","thirteen","fourteen","fifteen","sixteen","seve
nteen","eighteen","nineteen","twenty","twentyone","twentytwo")
DFT2.rmse.list[c("one","two","three","four","five","six")]=NULL
```

C. VBA SCRIPT FOR RUNNING HORIZON FORECAST OF NL OPTIMIZATION

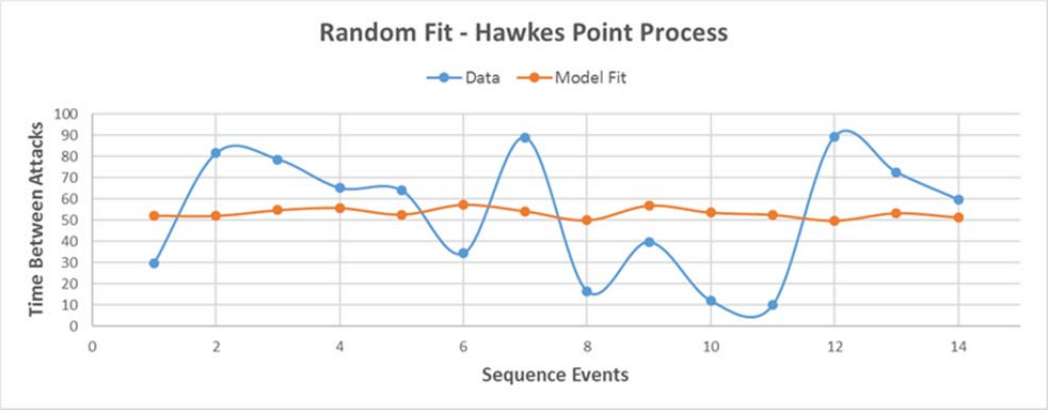
```
Sub Peturb()
        NL_Optim.Range("B2").Formula = Rnd() * 100
        NL_Optim.Range("B3").Formula = (Rnd() * ((-1) ^ Int(Rnd() * 10))) * 25
        NL_Optim.Range("B4").Formula = (Rnd() * ((-1) ^ Int(Rnd() * 10))) * 25
        NL_Optim.Range("B5").Formula = (Rnd() * ((-1) ^ Int(Rnd() * 10))) * 200
End Sub

Sub callGRG()
    Dim i, j As Integer
    Dim StartTime As Double
    Dim MinutesElapsed As String
    Dim A1, A2, W1, W2, P1, P2, O, bestSSE, rmse2 As Double
    Dim N As Integer
    StartTime = Timer
    NL_Optim.Range("H1") = 0
    N = Data.Range("B1", Data.Range("B1").End(xlDown)).Count
    For j = 0 To N - 14
        NL_Optim.Range("C11:C24").Value = Data.Range("A1:A14").Offset(j).Value
        NL_Optim.Range("H2") = j
        bestSSE = 1000000
        For i = 0 To 9
            Call Peturb
            SolverSolve UserFinish:=True, ShowRef:="ShowTrial"
            If IsNumeric(NL_Optim.Range("E7").Value) Then
                If NL_Optim.Range("E7").Value < bestSSE Then
                    A1 = NL_Optim.Range("B2")
                    W1 = NL_Optim.Range("B3")
                    P1 = NL_Optim.Range("B4")
                    O = NL_Optim.Range("B5")
                    bestSSE = NL_Optim.Range("E7")
                    rmse2 = NL_Optim.Range("G6")
                End If
            End If
        NL_Optim.Range("H1") = i
        Next i
        Result.Range("C2").Offset(j) = A1
```

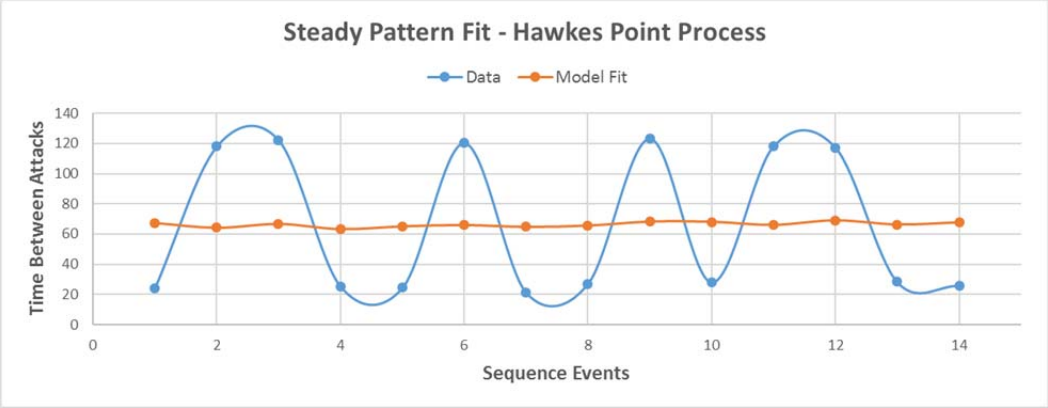
```
        Result.Range("D2").Offset(j) = W1
        Result.Range("E2").Offset(j) = P1
        Result.Range("F2").Offset(j) = O
        Result.Range("B2").Offset(j) = rmse2
        Result.Range("A2").Offset(j) = j + 13
    Next j
  MinutesElapsed = Format((Timer - StartTime) / 86400, "hh:mm:ss")
  MsgBox "This code ran successfully in " & MinutesElapsed & " minutes",
vbInformation
End Sub
```

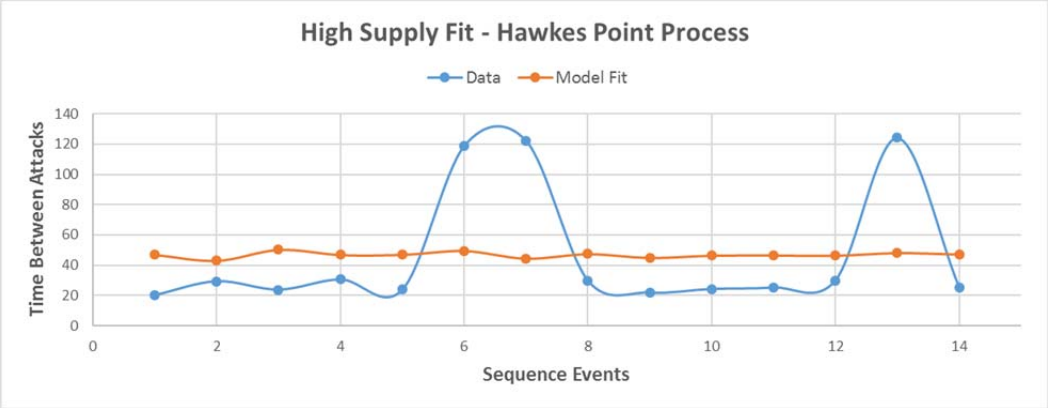
THIS PAGE INTENTIONALLY LEFT BLANK

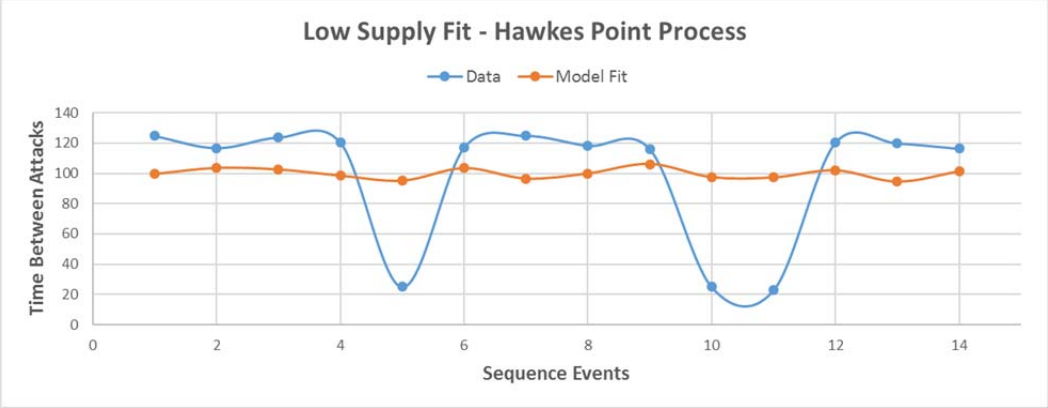
APPENDIX B. VISUALIZATION OF MODEL RESULTS

A. HAWKES POINT PROCESS

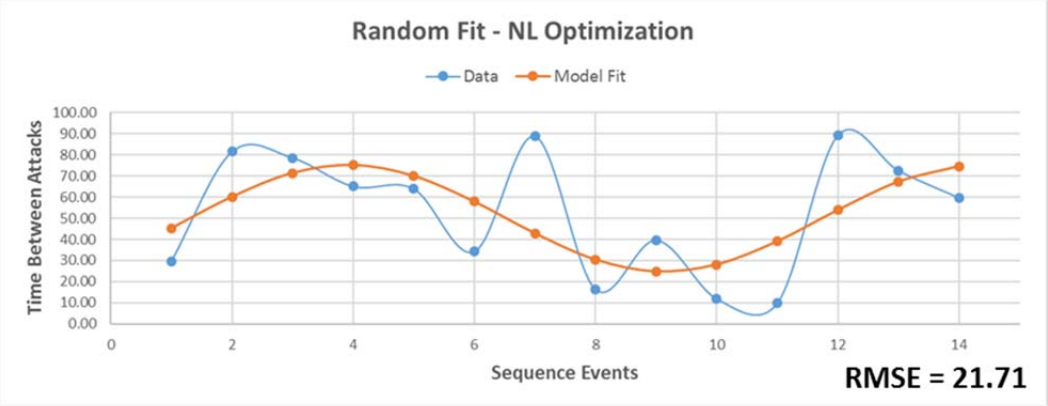


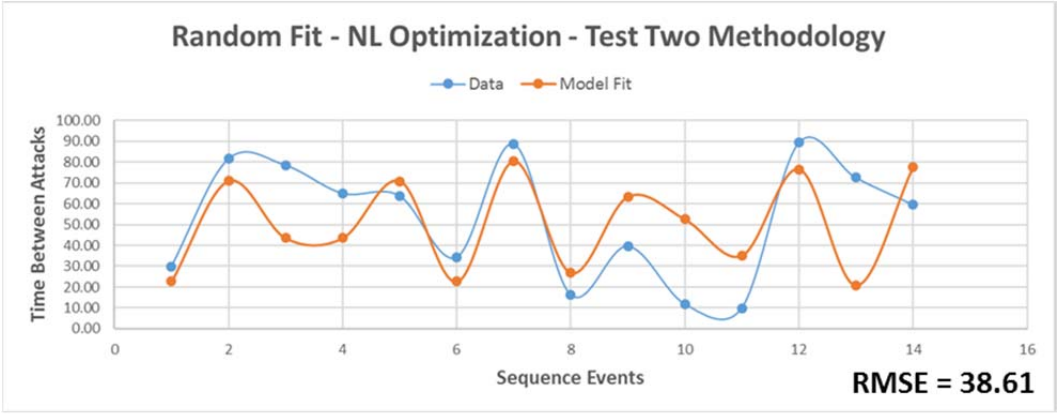


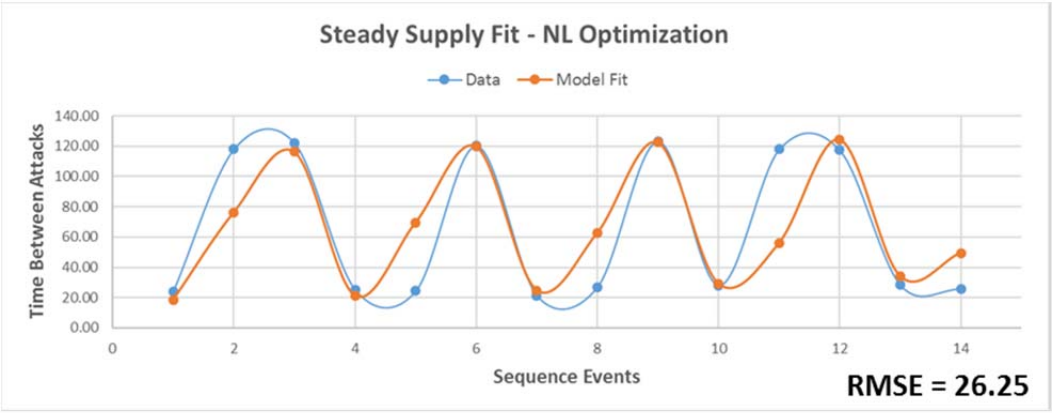


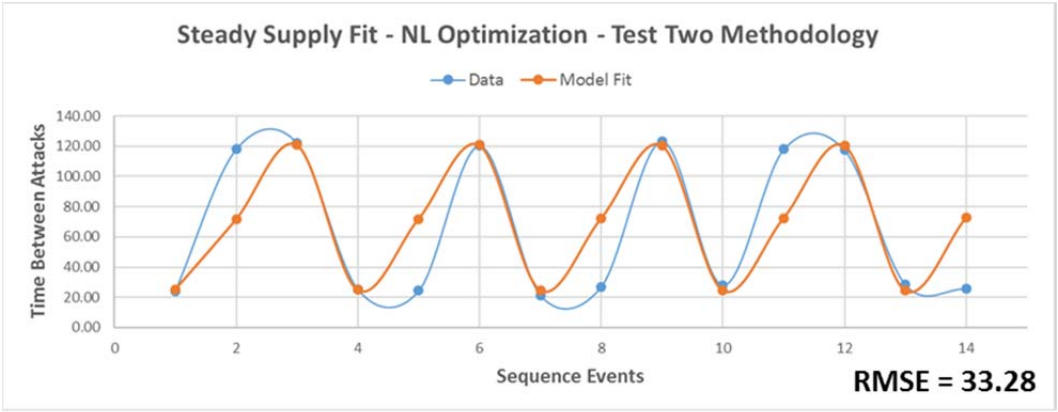


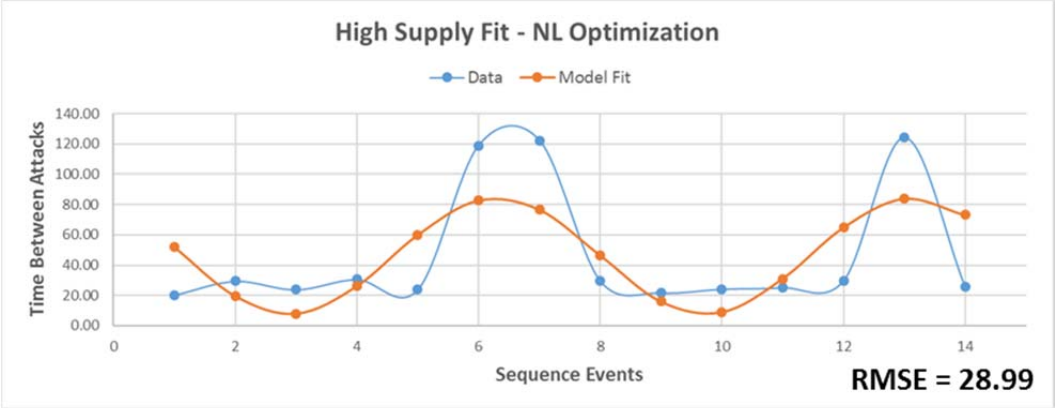
B. NL OPTIMIZATION

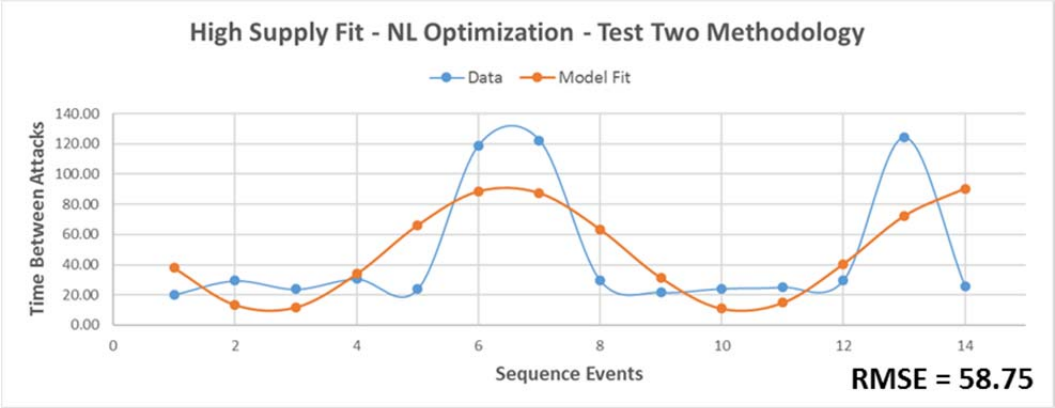


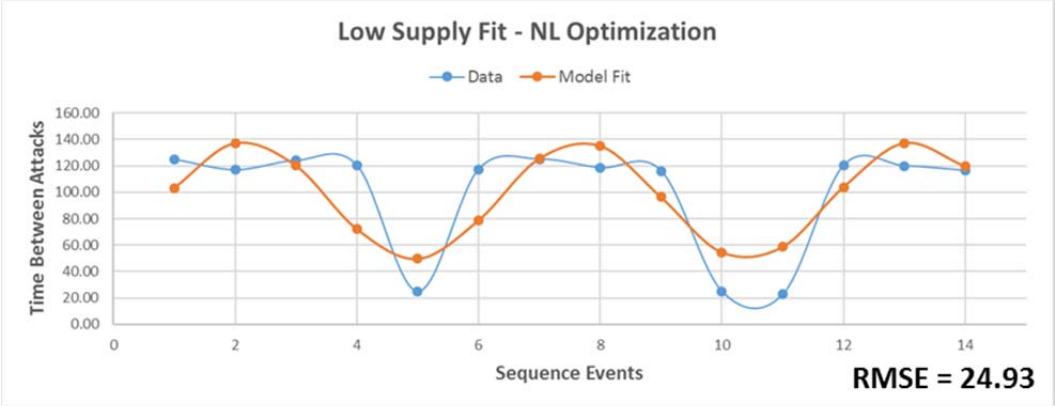

Random Fit - NL Optimization - Test Two Methodology
Data
Model Fit
Time Between Attacks
Sequence Events
RMSE = 38.61

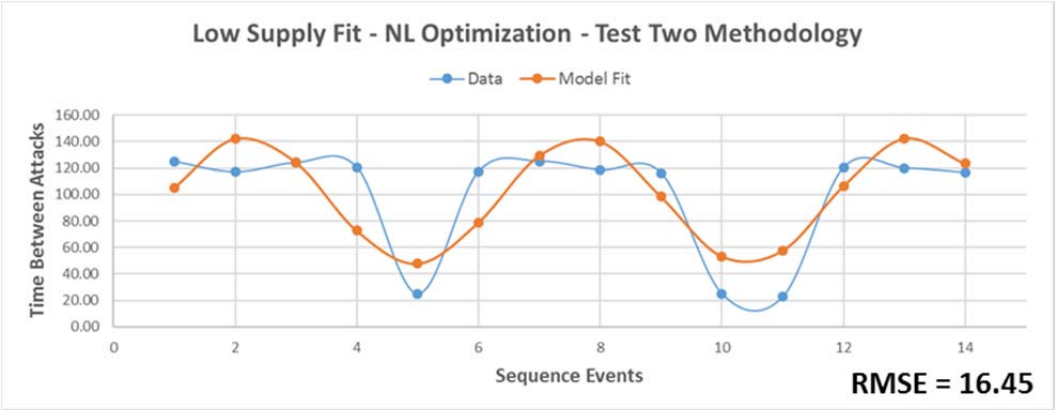

Steady Supply Fit - NL Optimization
Data
Model Fit
Time Between Attacks
Sequence Events
RMSE = 26.25


Steady Supply Fit - NL Optimization - Test Two Methodology
Data
Model Fit
Time Between Attacks
Sequence Events
RMSE = 33.28

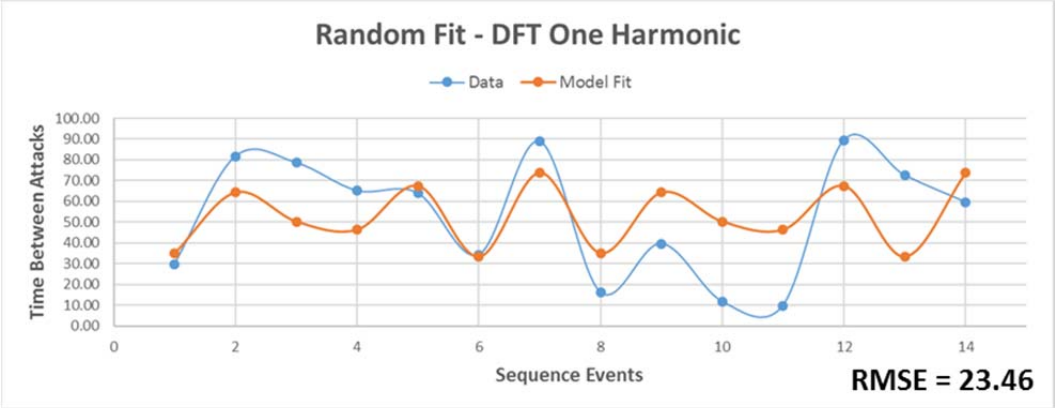

High Supply Fit - NL Optimization
Data
Model Fit
Time Between Attacks
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 28.99

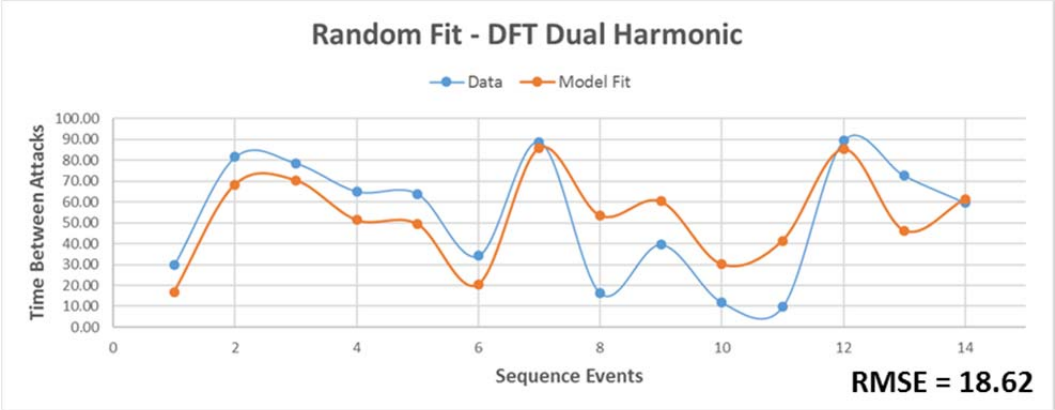

High Supply Fit - NL Optimization - Test Two Methodology
Data
Model Fit
Time Between Attacks
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 58.75

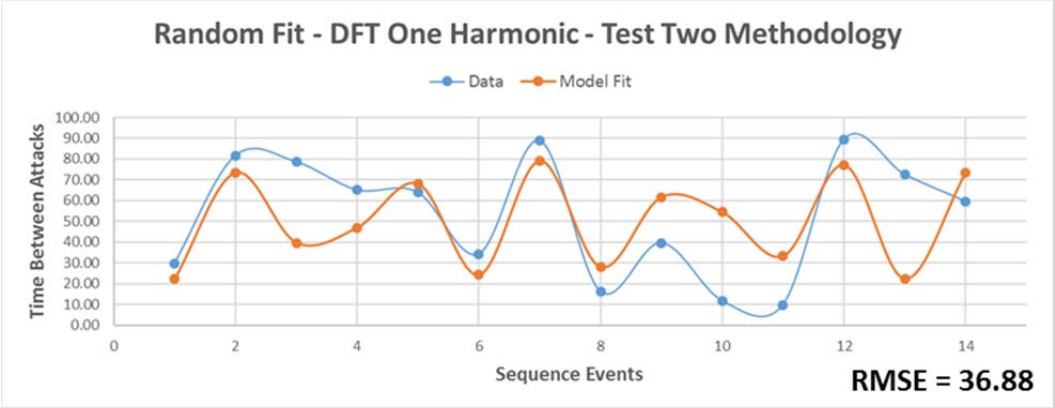

Low Supply Fit - NL Optimization
Data
Model Fit
Time Between Attacks
160.00
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 24.93

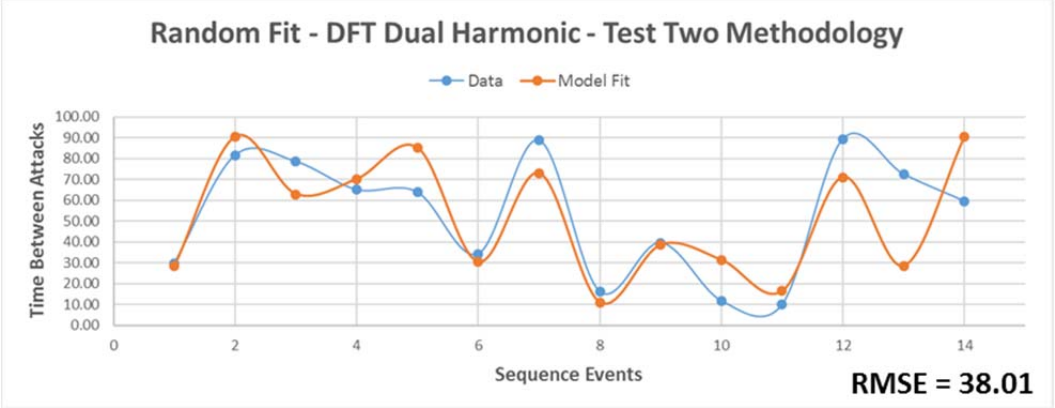


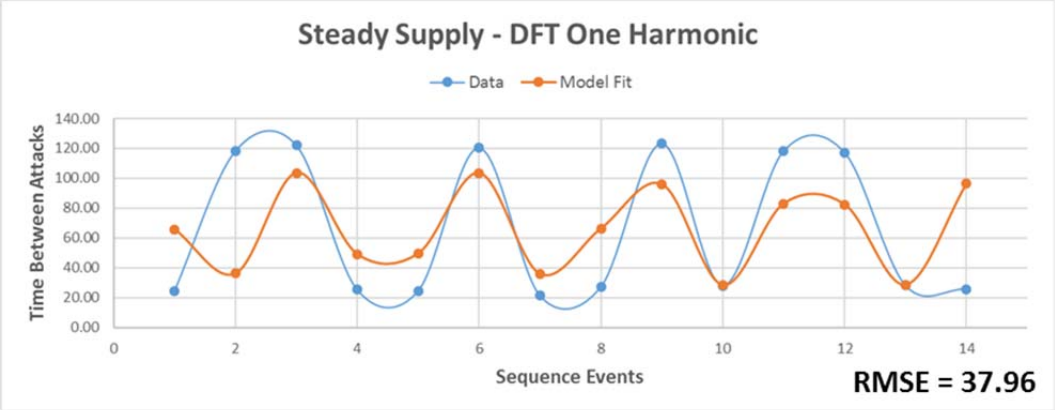
C. DISCRETE FOURIER TRANSFORMS

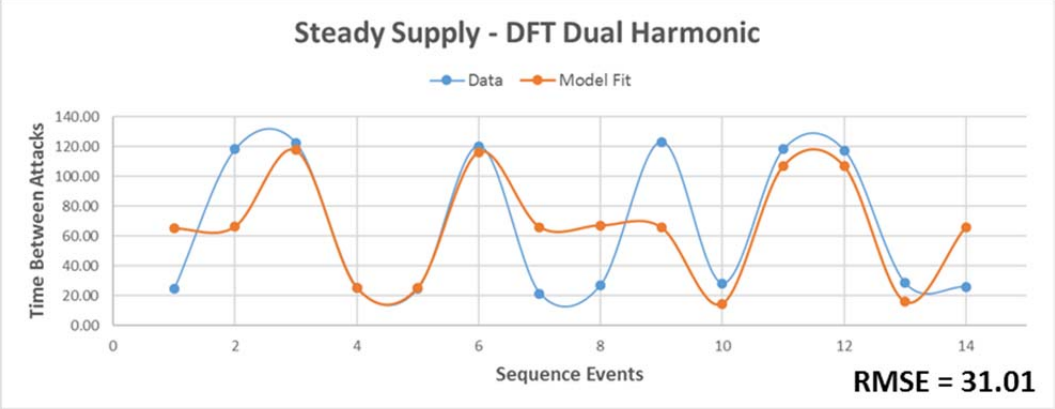


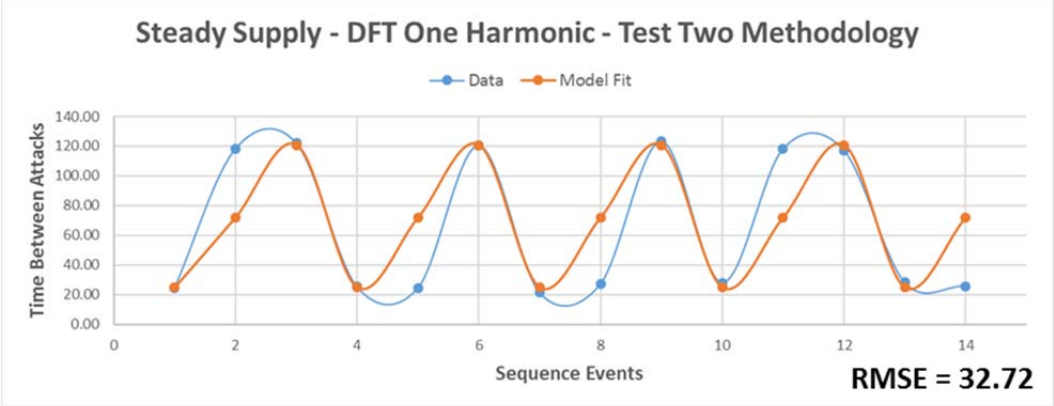

Random Fit - DFT Dual Harmonic
Data
Model Fit
Time Between Attacks
Sequence Events
RMSE = 18.62

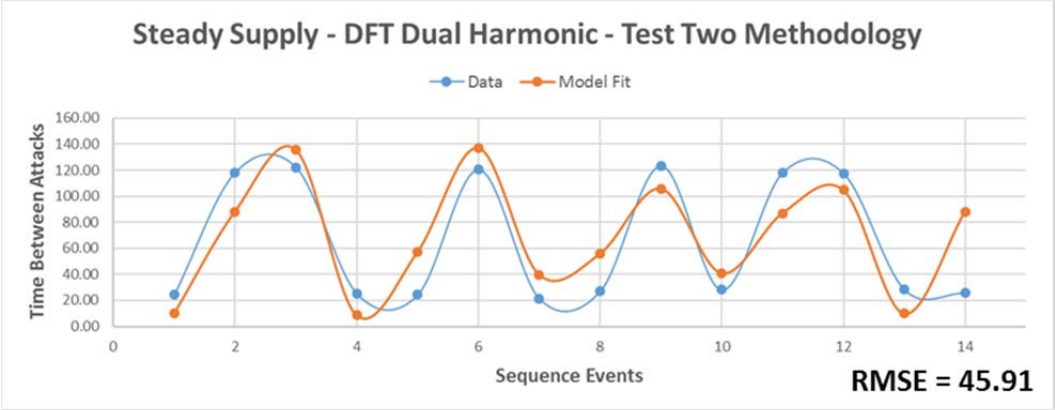

Random Fit - DFT One Harmonic - Test Two Methodology
Data
Model Fit
Time Between Attacks
Sequence Events
RMSE = 36.88

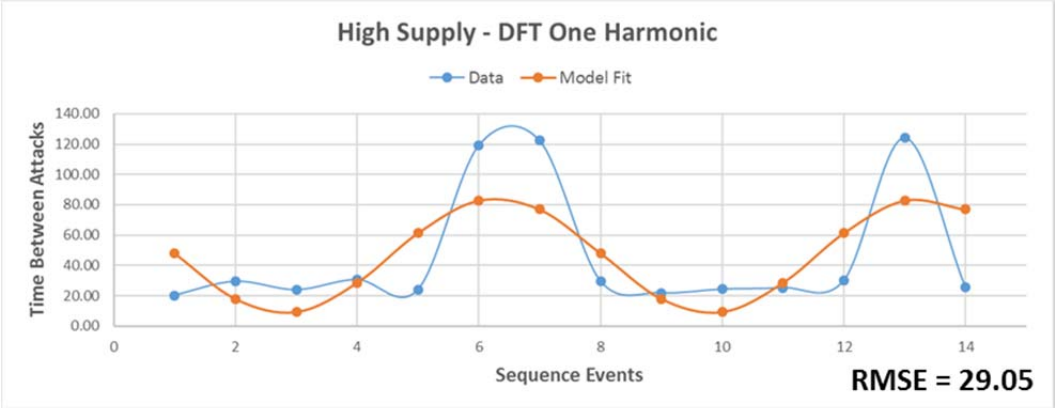

Random Fit - DFT Dual Harmonic - Test Two Methodology
Data
Model Fit
Time Between Attacks
Sequence Events
RMSE = 38.01

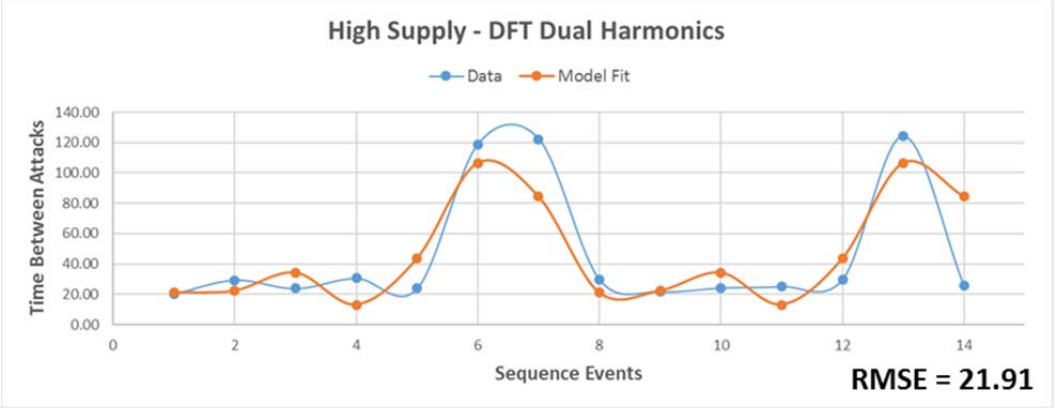

Steady Supply - DFT One Harmonic
Data
Model Fit
Time Between Attacks
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 37.96

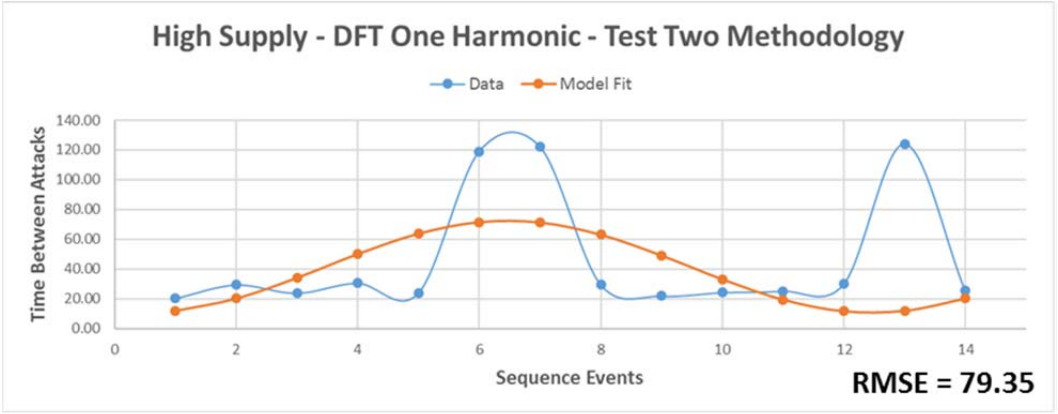

Steady Supply - DFT Dual Harmonic
Data
Model Fit
Time Between Attacks
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 31.01

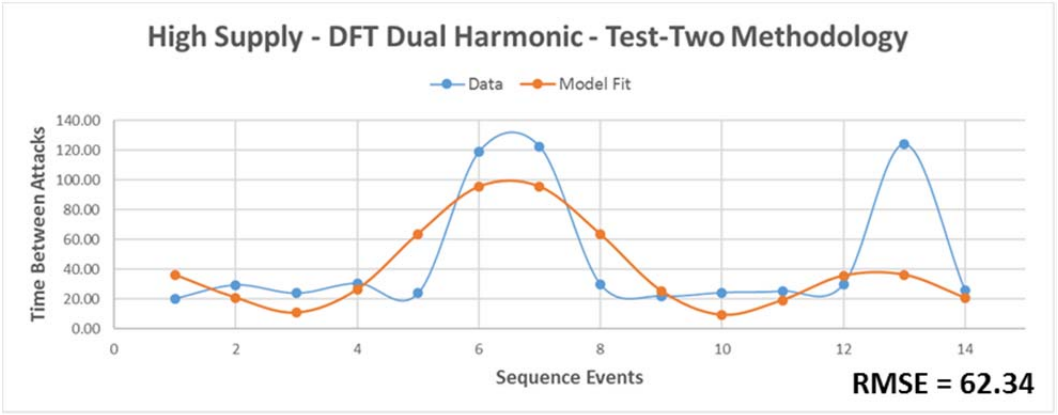

Steady Supply - DFT One Harmonic - Test Two Methodology
Data
Model Fit
Time Between Attacks
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 32.72

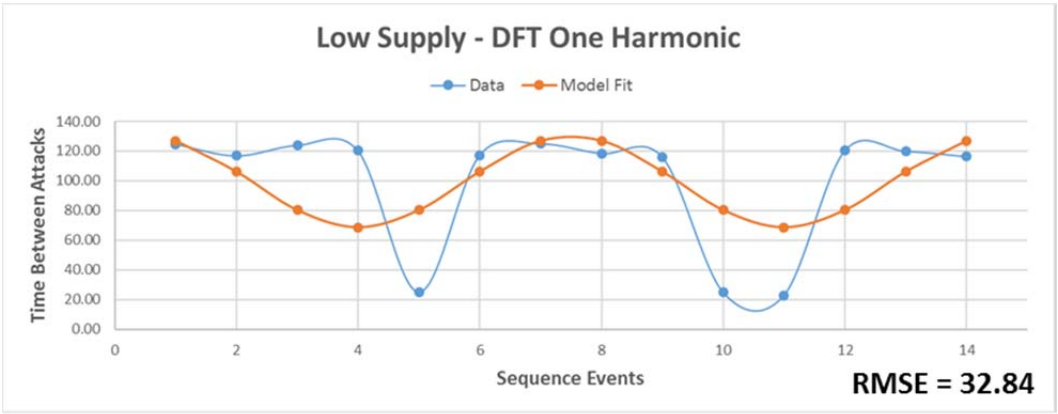

Steady Supply - DFT Dual Harmonic - Test Two Methodology
Data
Model Fit
Time Between Attacks
160.00
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 45.91

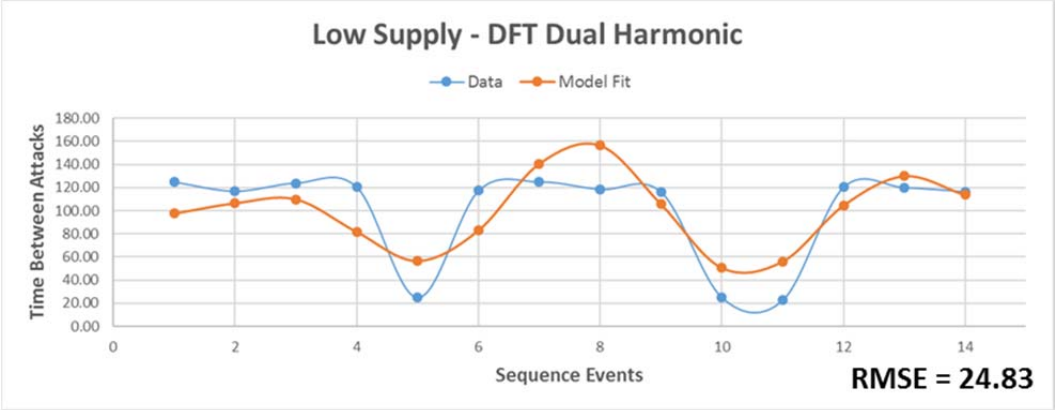

High Supply - DFT One Harmonic
Data
Model Fit
Time Between Attacks
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 29.05

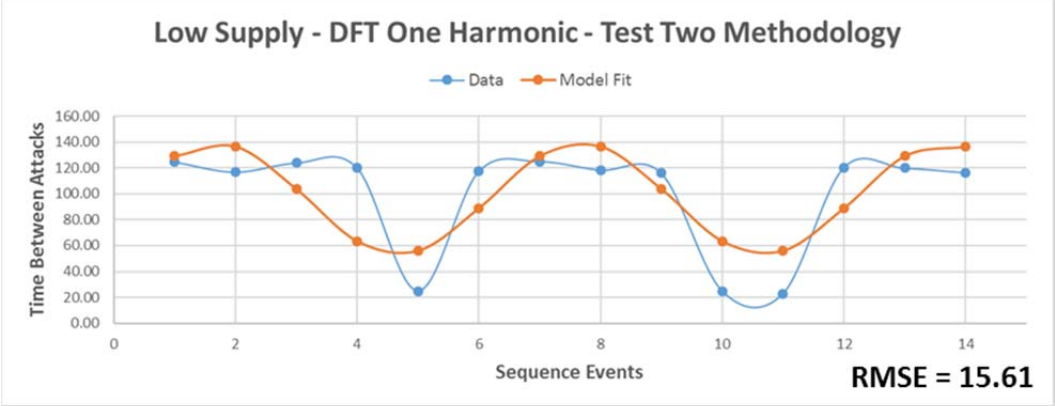

High Supply - DFT Dual Harmonics
Data
Model Fit
Time Between Attacks
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 21.91

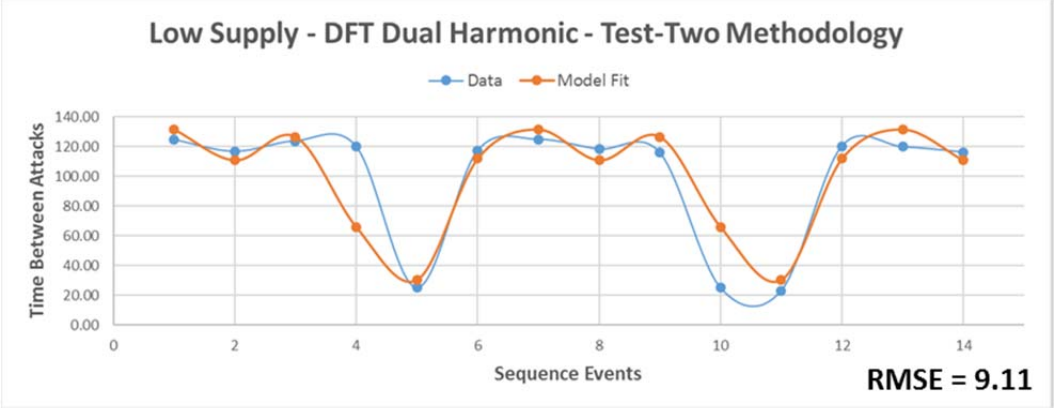

High Supply - DFT One Harmonic - Test Two Methodology
Data
Model Fit
Time Between Attacks
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 79.35


High Supply - DFT Dual Harmonic - Test-Two Methodology
Data
Model Fit
Time Between Attacks
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 62.34


Low Supply - DFT One Harmonic
Data
Model Fit
Time Between Attacks
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 32.84

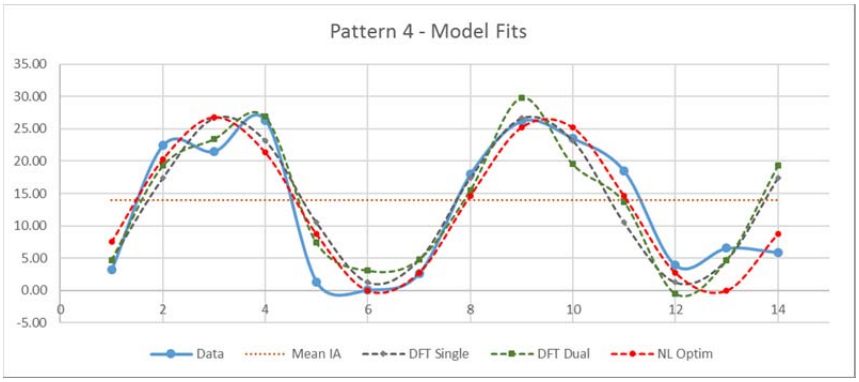

Low Supply - DFT Dual Harmonic
Data
Model Fit
Time Between Attacks
180.00
160.00
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 24.83


Low Supply - DFT One Harmonic - Test Two Methodology
Data
Model Fit
Time Between Attacks
160.00
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 15.61

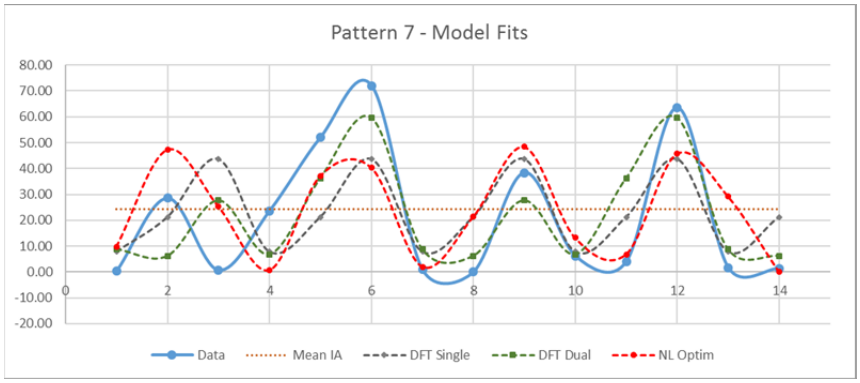

Low Supply - DFT Dual Harmonic - Test-Two Methodology
Data
Model Fit
Time Between Attacks
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
0
2
4
6
8
10
12
14
Sequence Events
RMSE = 9.11

APPENDIX C. EXAMPLE MODEL FITS OF CANDIDATE IRAQ PATTERNS

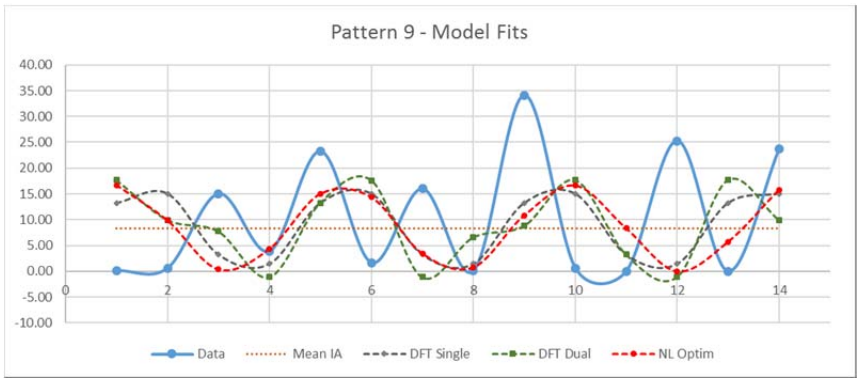
Pattern 4 - Data and Model Fits					
RMSE		7.75	8.23	9.55	5.06
Sequence	Data	Mean	DFT	DFT2	NL Optim
1	3.23	13.96	4.72	4.67	7.53
2	22.43	13.96	17.39	19.27	20.30
3	21.50	13.96	26.63	23.42	26.82
4	26.30	13.96	23.19	26.86	21.39
5	1.28	13.96	10.52	7.37	8.76
6	0.08	13.96	1.29	3.08	-0.01
7	2.67	13.96	4.72	4.78	2.78
8	18.00	13.96	17.39	15.51	14.68
9	26.17	13.96	26.63	29.83	25.25
10	23.50	13.96	23.19	19.52	25.25
11	18.43	13.96	10.52	13.67	14.66
12	3.88	13.96	1.29	-0.50	2.77
13	6.53	13.96	4.72	4.67	0.00
14	5.90	13.96	17.39	19.27	8.77



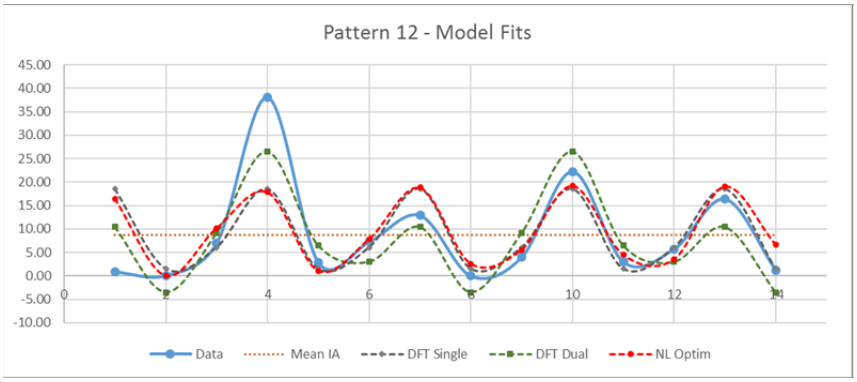
Pattern 7 - Data and Model Fits					
RMSE		22.66	14.62	6.02	19.59
Sequence	Data	Mean	DFT	DFT2	NL Optim
1	0.37	24.21	7.78	8.77	9.77
2	28.53	24.21	21.17	6.14	47.37
3	0.68	24.21	43.68	27.66	25.37
4	23.62	24.21	7.78	6.80	0.69
5	52.08	24.21	21.17	36.19	37.19
6	72.08	24.21	43.68	59.69	40.43
7	0.92	24.21	7.78	8.77	2.02
8	0.00	24.21	21.17	6.14	21.34
9	38.38	24.21	43.68	27.66	48.40
10	6.23	24.21	7.78	6.80	13.19
11	4.05	24.21	21.17	36.19	6.81
12	63.55	24.21	43.68	59.69	45.77
13	1.70	24.21	7.78	8.77	29.26
14	1.40	24.21	21.17	6.14	0.00



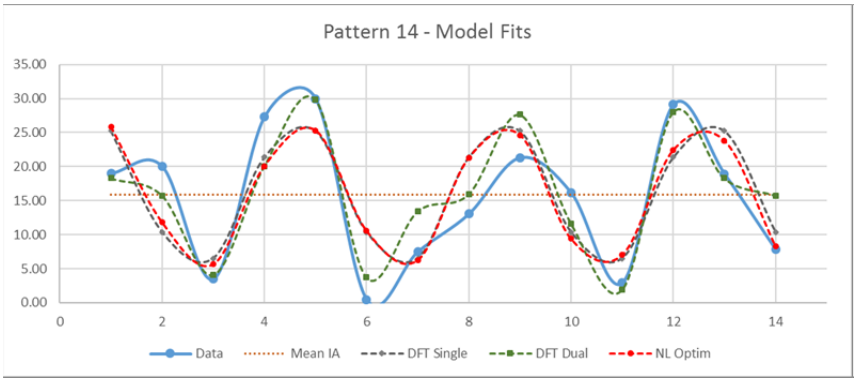
Pattern 9 - Data and Model Fits					
RMSE		11.99	4.44	7.98	2.02
Sequence	Data	Mean	DFT	DFT2	NL Optim
1	0.10	8.26	13.23	17.65	16.71
2	0.62	8.26	15.07	9.96	9.87
3	15.02	8.26	3.29	7.72	0.38
4	3.87	8.26	1.45	-1.12	4.30
5	23.23	8.26	13.23	13.24	15.00
6	1.67	8.26	15.07	17.61	14.37
7	15.98	8.26	3.29	-1.13	3.48
8	0.07	8.26	1.45	6.56	0.75
9	34.12	8.26	13.23	8.79	10.80
10	0.62	8.26	15.07	17.63	16.62
11	0.02	8.26	3.29	3.28	8.37
12	25.28	8.26	1.45	-1.10	0.00
13	0.00	8.26	13.23	17.65	5.68
14	23.70	8.26	15.07	9.96	15.80



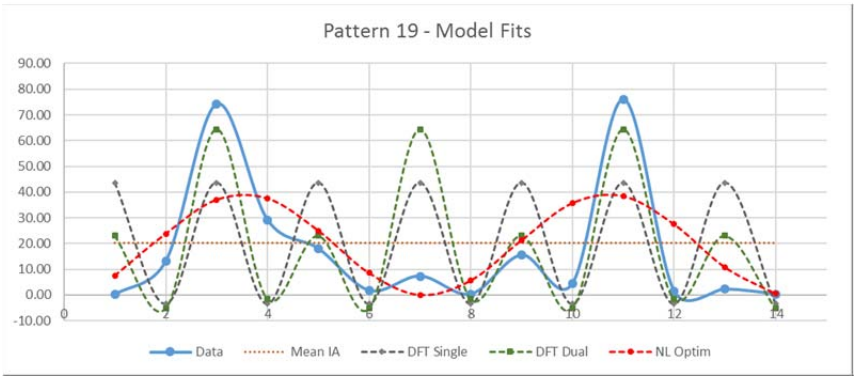
Pattern 12 - Data and Model Fits					
RMSE		7.57	1.50	5.40	4.26
Sequence	Data	Mean	DFT	DFT2	NL Optim
1	0.88	8.68	18.51	10.44	16.49
2	0.08	8.68	1.50	-3.51	-0.01
3	7.05	8.68	6.03	9.09	10.15
4	38.03	8.68	18.51	26.58	17.91
5	2.90	8.68	1.50	6.50	1.07
6	7.28	8.68	6.03	2.97	7.80
7	12.90	8.68	18.51	10.44	18.85
8	0.00	8.68	1.50	-3.51	2.61
9	4.03	8.68	6.03	9.09	5.54
10	22.22	8.68	18.51	26.58	19.25
11	3.00	8.68	1.50	6.50	4.52
12	5.75	8.68	6.03	2.97	3.49
13	16.40	8.68	18.51	10.44	19.11
14	1.27	8.68	1.50	-3.51	6.69



Pattern 14 - Data and Model Fits					
RMSE		6.08	4.83	5.63	3.48
Sequence	Data	Mean	DFT	DFT2	NL Optim
1	19.02	15.87	25.25	18.30	25.84
2	20.00	15.87	10.38	15.77	11.85
3	3.53	15.87	6.49	4.12	5.68
4	27.33	15.87	21.36	20.09	20.07
5	29.97	15.87	25.25	29.83	25.32
6	0.47	15.87	10.38	3.73	10.59
7	7.53	15.87	6.49	13.45	6.29
8	13.08	15.87	21.36	15.98	21.29
9	21.32	15.87	25.25	27.62	24.64
10	16.13	15.87	10.38	11.66	9.42
11	2.97	15.87	6.49	1.91	7.05
12	29.12	15.87	21.36	28.02	22.41
13	18.92	15.87	25.25	18.30	23.81
14	7.83	15.87	10.38	15.77	8.35



Pattern 19 - Data and Model Fits					
RMSE		18.83	12.78	16.36	6.08
Sequence	Data	Mean	DFT	DFT2	NL Optim
1	0.12	20.12	43.63	22.95	7.53
2	13.25	20.12	-3.38	-5.27	23.76
3	74.27	20.12	43.63	64.30	36.91
4	29.10	20.12	-3.38	-1.50	37.47
5	17.98	20.12	43.63	22.95	25.02
6	1.70	20.12	-3.38	-5.27	8.58
7	7.30	20.12	43.63	64.30	0.04
8	0.33	20.12	-3.38	-1.50	5.57
9	15.50	20.12	43.63	22.95	21.18
10	4.50	20.12	-3.38	-5.27	35.57
11	76.08	20.12	43.63	64.30	38.34
12	1.33	20.12	-3.38	-1.50	27.47
13	2.25	20.12	43.63	22.95	10.84
14	0.38	20.12	-3.38	-5.27	0.47



LIST OF REFERENCES

Ackerman S (2011) $265 Bomb, $300 billion war: The economics of the 9/11 era's signature weapon. *Wired* (September 08), https://www.wired.com/2011/09/ied-cost/.

AOAV (2014) The impact of IEDs: Three years of data, 2011–2013. *Action on Armed Violence* (December 1), https://aoav.org.uk/2014/three-years-ieds/.

Atkinson R (2007) Left of the boom: The struggle to defeat roadside bombs, Part 1. *The Washington Post* (September 30), http://www.washingtonpost.com/wp-dyn/content/article/2007/09/29/AR2007092900750.html?sid=ST2007092900754.

Bengini M, Furrer R (2012) Spatio-temporal improvised explosive device monitoring: Improving detection to minimize attacks. *Journal of Applied Statistics*. 39(11):2493–2508.

CIDNE—Intelligent Software Solutions. CIDNE—Operational Intelligence. Accessed February 12, 2016, http://www.issinc.com/government/products/cidne/.

Dunbar S (2005) Fitting a sine curve to data. Lecture, University of Nebraska-Lincoln, Lincoln, NE.

Fox E, Short M, Schoenberg F, Coronges K, Bertozzi A (2015) Modeling e-mail networks and inferring leadership using self-exciting point processes. *Journal of the American Statistical Association*. DOI:10.1080/01621459.2016.1135802

Hawkes A (1971) Spectra of some self-exciting and mutually exciting point processes. *Biometrika*. 58(1):83–90

Helms D (1999) *The Use of Dynamic Spatio-Temporal Analytical Techniques to Resolve Emergent Crime Series*. Report, Las Vegas Metropolitan Police Department, Las Vegas, NV.

Hyndman R, Koehler A (2006) Another look at measures of forecast accuracy. *International Journal of Forecasting*. 22(4):679-688

JIDA–Joint Improvised-Threat Defeat Agency. About JIDA. Accessed February 12, 2016, https://www.jieddo.mil/about.htm.

Koyak R (2009) Risk on Roads: A Modeling Approach, Part 1: Estimation models. Report. (Monterey, CA: Naval Postgraduate School).

Lewis E, Mohler G, Brantingham J, Bertozzi, A (2011) Self-exciting point process models of civilian deaths in Iraq. *Security Journal*. 25:244–264.

Marks C (2009) *Optimization-Based Routing and Scheduling of IED-Detection Assets in Contemporary Military Operations*. Massachusetts Institute of Technology, Cambridge, MA.

Masuda N, Takaguchi T, Nobuo S, Yano K (2013) Self-Exciting point process modeling of conversation event sequences. Temporal Networks. Springer-Verlag, Berlin (2013):245–264.

Mohler G, Short M, Brantingham P, Schoenberg F, Tita G (2012) Self-exciting point process modeling of crime. *Journal of the American Statistical Association*. 106(496):100–108.

Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*. 83(401):9–27.

Polat E (2007) *Spatio-Temporal Crime Prediction Model based on Analysis of Crime Clusters*. Middle East Technical University, Ankara, Turkey.

Shankar A (2014) Spatial and Temporal Modeling of IED Emplacements against Dismounted Patrols. *Mason Publishing*, George Mason University, Fairfax, VA.

Singer P (2012) The Evolution of Improvised Explosive Devices (IEDs). *Brookings* (February), http://www.brookings.edu/research/articles/2012/02/improvised-explosive-devices-singer.

Smith S (1997) *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Publishing, Poway, CA.

Stevenson M (2013) *Crime Analysis: The History and Development of a Discipline*. Western Oregon University, Monmouth, OR.

Toke I (2011) An Introduction to Hawkes Processes with Applications to Finance. Presentation, February 4, Ecole Centrale Paris, Châtenay-Malabry, France.

Washburn A (2006) *Continuous Network Interdiction*. Report. Naval Post-Graduate School, Monterey, CA.

Wilbanks M, Karsh E (2010) How the "Sons of Iraq" stabilized Iraq. *Middle East Quarterly* (Fall 2010):57–70.

Zoroya, G (2013) How the IED changed the U.S. military. *USA Today* (December 19), http://www.usatoday.com/story/news/nation/2013/12/18/ied-10-years-blast-wounds-amputations/3803017/.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California